UNIVERSITY OF LONDON THESIS

Degree: PhD
Year: 2006
Name of Author: TANG, J. C. Y.

When Does Size Matter?

The Effects of Task Relevance in the Processing of Numerical Magnitude

Joey Chung Yee Tang

Institute of Cognitive Neuroscience
University College London

PhD Psychology

Abstract

The present thesis examines the processing of numerical magnitude using a series of Stroop variants; existing paradigms are modified and new ones devised. The Stroop effect, observed when numerical magnitude is the task-irrelevant dimension, has been widely used as an indicator for autonomous processing. However, it is argued that this effect does not provide a sensitive enough measure for the degree of information processing. Instead, current findings have demonstrated that the reversed numerical distance effect observed during physical size comparison of digits (when numerical magnitude is the task-irrelevant dimension) can be used as a reliable measure for *refined* autonomous processing. Neuroimaging data are consistent with this proposal. Others factors which influence numerical magnitude processing, such as effects of writing system and familiarity, are also examined. Findings are discussed with respect to existing theoretical accounts. In addition to the relative strength and speed of processing, stimulus discriminability and familiarity also contribute to a more comprehensive understanding of number comparison.

The autonomous property of numerical magnitude processing during parity judgement and numerosity matching is also investigated. Current findings indicate that numerical magnitude processing is, in most cases, refined. Experimental designs have significant contributions as to whether refined processing can be elicited. Moreover, the present thesis provides evidence supporting the ideas that (1) representations of numerical magnitude follow the principle of cardinality and are therefore linear, and (2) numbers evoke discrete representations which are distinct from continuous presentations evoked by other quantitative dimensions such as physical size.

Future research directions are suggested, with an emphasis on non-symbolic stimuli which can potentially be used as diagnostic tools for dyscalculia with young children, on the assumption that the aetiology of this developmental condition rests on a more general quantity processing deficit.

Acknowledgments

I extend my sincere gratitude and appreciation to all those who made this thesis possible. I am deeply indebted to my supervisors Professor Brian Butterworth and Dr Daniel Glaser whose stimulating suggestions and encouragement helped me in all the time of my research. Many thanks go to Professor Ruth Stavy from University of Tel Aviv who offered insightful comments on my thesis. I would also like to thank my examiners, Dr Elizabeth Isaacs from the Institute of Child Health and Dr Wim Fias from Ghent University, for their care in evaluating my work.

I would like to acknowledge with much appreciation the support from Nippon Telegraph and Telephone (NTT) Corporation. Many thanks are due to Dr Tatsuya Hirahara for allowing me to carry out substantial work at the Human and Information Science Laboratory. Many thanks go to the staff who spared their time out of tight schedules to assist my research during my stay.

I would also like to thank my friends for their help and interests. Special thanks go to Ms Raffaella Moro, Ms Eva-Maria Ebner, and Dr Marinella Cappelletti for their friendship and support, and to Mr Wing To for his patience and love.

Finally, I would like to give special thanks to my parents whose unconditional love enabled me to complete this work.

Table of Contents

Table of Tables

Table of Figures

"Take from all things their number and all shall perish. Take calculation from the world and all is enveloped in dark ignorance..."

Isidore of Seville (ca 600 AD)

Preface

Isidore's infamous quote (in Crosby, 1997) undoubtedly captures the importance of numbers. Indeed, numbers[1] are used in countless everyday activities and in different contexts and, accordingly, they convey different meanings (Fuson, 1992). Numbers represent physical dimensions such as height and weight, temporal dimensions such as time and age, and other meaningful dimensions such as price and temperature. Numbers are also used in ordering things in our environment, such as pages in a book and houses in a street. Furthermore, numbers constitute the far most preferred coding system to identify different exemplars in a given category, e.g., telephone users, bus lines, and books (which are identified by the International Standard Book Number or ISBN). More importantly, numbers enable more complex mental manipulations since they are the building blocks for calculations. In this context, they refer to specific numerosities which may be transformed through arithmetical operations (e.g., 3 + 5 = 8). Given that numbers play a major role in how we understand, manipulate and interact with our environment, concerted effort in research is required to deepen the understanding of the numerical processing[2].

Our understanding of the mechanisms involved in numerical processing comes from the following major sources: behavioural performance of normal adults (e.g., Banks, 1977; Besner & Coltheart, 1979; Dehaene, Bossini, & Giraux, 1993; Foltz, Poltrock, & Potts, 1984; Henik & Tzelgov, 1982; Moyer & Landauer, 1967; Paivio, 1975), neuroimaging studies on normal adults (e.g., Pinel, Dehaene, Rivière, & Le Bihan, 2001; Pinel, Piazza, Le Bihan, & Dehaene, 2004),

[1]The term *number*, in the present thesis, refers to any symbol, whatever its notational code, representing a number, whereas the term *Arabic numeral* refers to one of the following: in digit form, e.g., 3; in word form (*verbal numeral*), e.g., three, which may be spoken (*spoken verbal numeral*) or written (*written verbal numeral*).

[2]Numerical processing refers to any cognitive manipulation of numbers or other numerical stimuli, e.g., a set of dots.

developmental studies (e.g., Fuson, 1988; Gelman & Gallistel, 1978; Girelli, Lucangeli, & Butterworth, 2000), performance of individuals who have suffered brain damage (e.g., Butterworth, Cappelletti, & Kopelman, 2001; Cipolotti & Butterworth, 1995; Cohen, Dehaene, & Verstichel, 1994; Deloche & Seron, 1987; McCloskey, Sokol; & Goodman, 1986; Rossetti, Jacquin-Courtois, Rode, Ota, Michel, & Boisson, 2004; Vuilleumier, Ortigue, & Brugger, 2004), and computational modelling of numerical representation and processing (e.g., Zorzi & Butterworth, 1999). Although these disciplines employ different methodologies, findings and inferences from these above disciplines should not be considered separately; they together build a more comprehensive picture of the cognitive architecture of numerical processing. For example, the developmental sequence of numerical competence has benefited the understanding of numerical processing in normal adults (e.g., Girelli et al., 2000), and the detailed analysis of impaired performance patterns in brain-lesioned patients sheds light onto the cognitive architecture and functioning of normal numerical abilities (e.g., Caramazza, 1986; McCloskey & Caramazza, 1988). Furthermore, computer simulations have allowed models to be "lesioned" producing deficits which match human neuropsychological data (e.g., Lories, Aubrun, & Seron, 1994), and hence providing further insights into the cognitive architecture and functioning of normal numerical abilities.

The current thesis focuses on fundamental numerical processing, hence normal subjects are considered. Here, the processing of numerical magnitude is investigated with respect to task relevance. It has been suggested that the mere presence of an Arabic numeral may determine the activation of its magnitude representation (e.g., Dehaene et al., 1993; Henik & Tzelgov, 1982), and research within the past 30 years has supported the proposal (e.g., Besner & Coltheart, 1979; Girelli et al., 2000; Henik & Tzelgov, 1982). In addition, the mere presence of verbal numerals has also been observed to elicit similar activation of magnitude representation (e.g., Foltz et al., 1984). However, existing methodologies are coarse, thus failing to capture precisely the degree of information processing during task-relevant and -irrelevant conditions. In the present work, new experimental paradigms are devised in order to probe numerical magnitude processing under task-irrelevant conditions.

1 Literature Review

The present chapter provides an overview of information conveyed by numbers and clarification of terms used in number research. It highlights the fundamental issue of what numerical magnitudes mean, or more specifically, what representations are evoked by numbers. Literature on investigation of numerical magnitude processing is reviewed, with respect to existing theoretical frameworks. Existing methodologies are also examined.

1.1 Overview on Numerical Information and Numerical Representation

Numbers convey different meanings in different contexts (Fuson, 1992). The fundamental question of the current thesis is what information numbers readily elicit under both task-relevant and -irrelevant conditions. This question can be broken down into: What do numbers stand for? What information do numbers carry? How do we mentally represent numbers? Do we process numbers in an autonomous[3] fashion? Which properties of numbers do we process?

One of the first pieces of numerical information we learn as children is ordinality – the order of the number sequence: one, two, three, four, five, six, seven, eight, nine, ten, and so on. It is important to note that this piece of information does not refer to numerical magnitude (size) or, more precisely, numerosity (the number of items) per se, but merely the order of the verbal items. Children of two or three years often think of the first few number words as just one big world – "onetwothreefourfive" – and it takes them some time to learn that this big word

[3]The terms "automatic" and "autonomous" have been used to refer to a process that takes place even when it is irrelevant to the task at hand. However, when clearly stated, the access to numerical information has been said to be automatic by virtue of being autonomous, i.e., it begins and runs to completion without intention (Zbrodoff & Logan, 1986). Autonomy, therefore, by definition, is a property of automatic processes, i.e., processes which are fast, effortless, autonomous and unconscious (e.g., Logan, 1980; Posner & Snyder, 1975; Shiffrin & Shneider, 1977).

is really five small words (Fuson, 1992).

Research with children (e.g., Wynn, 1990) clearly indicates the dissociation between ordinality (sequence information) and cardinality (numerosity information). Cardinality, according to Zorzi and Butterworth (1999), implies that "smaller numbers denote proper subsets of the sets denoted by bigger numbers", thus a three consists of a two and a one, where a two consists of two ones. In Wynn's (1990) study, children were classified into "counters" (those who give you reliably the last word of the count to indicate numerosity) and "grabbers" (those who have not yet grasped the role of numbers in counting). The following conversation shows how a grabber, Adam, lacks full understanding of the cardinality principle:

> EXPERIMENTER (E) So how many are there?
> ADAM (A) (Counting three objects...) One, two, five!
> E (Pointing towards the three items): So there's five here?
> A No, that's *five* (pointing to the item he'd tagged "five")
> ...
> E What if you counted this way, one, two five? (Experimenter counts the objects in a different order than Adam has been doing)
> A No, that's five (pointing to the one he has consistently tagged "five")

In Adam's case (Wynn, 1990), he believed that each number word was just a label attached to an object and that one particular object, thus, although he was showing some understanding to the concept of ordinality (i.e., assigning a number word to each item in a sequential manner, despite the fact that he used one wrong number word), he failed to understand the essence of cardinality – that counting gives you the numerosity of the set. According to Gelman and Gallistel (1986), children of three-and-a-half are proficient at giving the last word in the count as the number of objects counted, hence they have acquired the "cardinal word principle" (see also Gelman & Meck, 1983).

Numbers have different values, or sizes, or more precisely, numerical sizes; however, to avoid confusion from now on, the term "numerical magnitudes", or simply "magnitudes[4]" will be used to refer to the meaning of numbers, whereas "sizes", or more precisely, "physical sizes" refers to physical dimension. What do magnitudes mean? Numbers, like physical dimensions such as volume, can be large or small, but are numbers mentally represented as a small-to-large continuum similar to a physical size continuum or do they refer to discrete numerosities?

There has been debate as to how numbers are actually represented; some researchers believe that numbers are represented in a similar way as other analogue (or continuous) dimensions such as length, and that they are processed in similar fashion, for example, according to Moyer and Landauer's (1967), "the decision process... is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line" (see also Dehaene, 1992; Dehaene & Cohen, 1995; McCloskey, 1992). Others, however, have argued that numbers evoke discrete representations of numerosity (e.g., Zorzi & Butterworth, 1999; see also Castelli, Glaser, & Butterworth, 2006). According to Zorzi and Butterworth (1999), "we think of a whole number as denoting the numerosity (or cardinality) of a set with discrete members", arguing that "analogue representations, however, fail to capture our intuitive notion of whole numbers, and whole-number arithmetic." Indeed, recent findings from Castelli et al.'s (2006) support the distinction between analogue (continuous) and discrete representations.

Closely related to the concept of ordinality, research has suggested numbers are spatially organized and have been described as a "mental number line" along

[4]The term "magnitude", according to Zorzi and Butterworth (1999), refers to the numerosity a number stands for. However, it is not the only semantic dimension of numbers, others include parity status.

which numbers are positioned from the left to the right side (Dehaene et al., 1993; Restle, 1970; Seron, Pesenti, Noël, Deloche, & Cornet, 1992. Note that, according to these authors, the mental number line implies an analogue representation of numbers.) According to Seron et al. (1992), "some people declare that they possess a personal visual representation of numbers: some automatically "see" the numbers they are confronted with in a precise location in a structured mental space, others "associate" specific colours with given numbers", echoing early research by Galton (1880) who reported that some people "saw" number lines with strange shapes, which he called "number forms". Dehaene et al.'s (1993) findings, from a series of nine experiments, strongly support the idea that numbers are organized in space; it was observed that "large numbers preferentially elicited a rightward response, and small numbers a leftward response" and this effect has been named the Spatial-Numerical Association of Response Codes (SNARC) effect (see Section 1.2.2). This effect was found to be linked to the direction of writing, "as it faded or even reversed" in right-to-left writing Iranian subjects (Dehaene et al., 1993).

Interestingly, behavioural performance of neglect patients in number tasks provides support for a distinction between spatial and semantic (magnitude) representations of numbers. In a number comparison task, lateralized disturbance in the mental representation of numbers was observed in neglect patients – "when asked to judge whether a single number shown at fixation was smaller or larger than "5", patients with neglect were selectively slower to respond to "4", but when asked to compare numbers to "7" they were selectively slower to respond to "6"." (Vuilleumier et al., 2004). In line with this finding, Zorzi, Priftis, & Umiltà (2002) reported that mental bisection between two numbers in neglect patients was systematically shifted to the right, i.e., towards larger numbers. This pattern of behaviour is in parallel with patients' bias when asked to mark the centre of a physical line (e.g., Marshall & Halligan, 1989). The impaired performance is consistent with the concept that mental number line is implicitly used in representing numbers from left to right, and when the representation of the left visual field is impaired as in neglect patients, lateralised disturbance in performance results. However, in a number Stroop task where neglect patients had to judge whether a single digit number was physically large

(~ 4° of visual angle) or small (~ 1.5° of visual angle) while ignoring the numerical magnitude of the number, they did not show a selective loss of the Stroop effect with leftward numbers i.e., below 5 (Vuilleumier et al., 2004). Instead, these patients showed an intact number Stroop effect – numbers both larger and smaller than 5 elicited interference in physical size comparisons. This finding suggests that the patients had intact semantic representations of numbers which could be dissociated from their impaired spatial representations of numbers, supporting the distinction between semantic (magnitude) and spatial representations of numbers.

As well as numerical magnitudes, numbers convey other information, including learnt facts such as parity (oddness and evenness), prime (number which is only divisible by 1 and itself), and arithmetical facts (addition, subtraction, multiplication, and division). Research has suggested that some of the above properties of numbers may be readily available even under task-irrelevant conditions. For example, using a number-matching paradigm in which mental arithmetic was task-irrelevant (LeFevre, Bisanz, & Mrkonjic, 1988), Galfano, Rusconi, and Umiltà (2003) provided evidence for automatic activation of multiplication facts. In Galfano et al.'s (2003) experiment, subjects were presented with a cue of two digits followed by a digit probe and had to decide whether any of the digits in the cue matched any of those in the probe. Subjects were slower in responding "no" to probes that were numbers adjacent to the product in the multiplication table related to the cue digits than to probes that were unrelated to the cue. Such a finding suggests that subjects automatically activated not only the product of the cue digits, but also the neighbours of the product. This is consistent with the idea that multiplication facts are stored in a highly related network in which activation spreads from the product node to adjacent nodes.

The above summary gives some idea of the wide scope of number research. The present thesis discusses ways to study numerical information processing, focusing on the processing of numerical magnitude under both task-relevant and -irrelevant conditions. Existing methodologies are critically discussed and new experimental paradigms have been devised and examined. Both behavioural and neuroimaging techniques are used, with aims to gain knowledge in numerical processing and numerical representations.

1.2 Markers of Semantic Number Processing

As mentioned before, numbers are used in numerous different contexts and convey, accordingly, different meanings, only some of which are truly numerical (e.g., cardinality/ numerosity). Nevertheless, magnitude[5] is undeniably the most distinctive attribute of a number. So far, two different effects have been used as markers of semantic number processing, namely, the numerical distance effect and the SNARC effect.

1.2.1 The Numerical Distance Effect

The numerical distance effect refers to the inverse relation between the time required to compare two numbers and the numerical distance (or difference) between them (e.g., Banks, Fujii, & Kayra-Stuart, 1976; Duncan & MacFarland, 1980; Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Foltz et al., 1984; Hinrichs, Yurko & Hu, 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977). For example, it takes longer to select the larger (or smaller) number in the pair 2 and 3 than in the pair 2 and 9. This effect has also been observed with 2-digit numbers (e.g., Dehaene, 1989; Dehaene Dupoux, & Mehler, 1990; Hinrichs et al., 1981). Moreover, research suggests that the distance effect is resistant to extensive practice (Dehaene et al., 1990; Fairbank, 1969).

The distance effect is not restricted to the numerical dimension, but occurs in other comparative judgements, e.g., physical size of numbers (Pinel et al., 2004),

[5]The term "magnitude" here refers to the numerosity a number stands for (Zorzi & Butterworth, 1999), and does not imply any reference to the analogue magnitude representation discussed by other authors (e.g., Dehaene, 1992; Dehaene & Cohen, 1995; McCloskey, 1992; Moyer & Landauer, 1967).

ordering of alphabets (Lovelace & Snodgrass, 1971; Parkman, 1971), and imaginary/ memorised representations such as angle size between the hour and the minute hands of an imaginary clock (Paivio, 1978) and memory of animal size (Moyer, 1973).

The discoverers of the numerical distance effect, Moyer and Landauer (1967), stated that "the decision process... is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line." Dehaene and colleagues, in a series of papers, argued that numbers are converted to analogue representations in the form of a mental number line (Dehaene, 1992; Dehaene & Changeux, 1993; Dehaene et al., 1990; see also McCloskey, 1992; Restle, 1970), and that this number line is said to be compressive (i.e., non-linear), obeying the Weber-Fechner logarithmic law (Weber, 1834; Fechner, 1860). Accordingly, the subjective difference between two numbers depends on their positions on the number line, so that the subjective difference (Δ) between the number N and N + 1 decreases as N increases. This can be expressed by the equation:

$$\Delta = (N + 1) / N$$

However, the concept of analogue representations for numbers has been challenged by Zorzi and Butterworth (1999), who argued that "Analogue representations, however, fail to capture our intuitive notion of whole numbers". Instead, these authors proposed that "number representations are ordered by numerosity: smaller numbers denote proper subsets of the sets denoted by bigger numbers" and crucially, that numerals are mapped linearly on to magnitude representations (discrete numerosities). Through the use of computational modelling, Zorzi and Butterworth (1999) observed the numerical distance effect in number comparison and found that the Welford function[6] (Welford, 1960)

[6]The Welford function (Welford, 1960) states that $RT = a + k \log [L / (L - S)]$ where L and S are the larger and smaller magnitudes respectively, and a and k are constants.

accounts for 88.3% of the variance of the model's reaction times. The same equation accounts for 50% of the variance of human data in number comparison (e.g., Dehaene, 1989; Moyer & Landauer, 1973). According to the Welford function (Welford, 1960), subjective distance (Δ) between the number N and N + 1 can be expressed by the following equation:

$$\Delta = \log \left(\frac{N+1}{N} + k \right)$$

where k is a constant.

Within Zorzi and Butterworth's (1999) framework, numerical representations are assumed to be linear, and the subject distance (Δ) predicted by the Welford function (Welford, 1960) does not decrease/ increase with numerical magnitudes. The authors concluded that the compressive effect on comparison times should not be attributed to the representation of numerical magnitudes; instead, they argued that the effect "emerges by virtue of the non-linear interactions that are intrinsic to the decision process itself".

Both analogue and discrete representation views provide support for the numerical distance effect as a reliable index of semantic information processing – the former suggests a comparison between analogue magnitudes on a mental number line, whereas the latter suggests a comparison between numerosity representations. However, as suggested by Dehaene and Akhavein (1995), one cannot exclude the possibility that the distance effect may partially arise from the intralexical association between consecutive or close numbers in the counting sequence. According to this hypothesis, the effect would reflect (at least partially) lexical distance rather than purely semantic distance. In Dehaene and Akhavein's (1995) number comparison experiment, lexical distance was kept constant while numerical distance was varied. The numerically close pairs used were 9 13, 13 9, 18 20, and 20 18, whereas numerically far pairs were 3 19, 19 3, 12 80, and 80 12. The constant lexical distance was constructed by the difference in stack positions (see Table 1.1). While controlling lexical distance, Dehaene and Akhavein (1995) still observed the numerical distance effect, thus supporting the semantic interpretation of the effect.

Table 1.1 The construction of a constant lexical distance in Dehaene and Akhavein (1995)'s experiment

	Ones	Teens	Tens
0	zero	ten	
1	one	eleven	
2	two	twelve	twenty
3	three	thirteen	thirty
4	four	fourteen	forty
5	five	fifteen	fifty
6	six	sixteen	sixty
7	seven	seventeen	seventy
8	eight	eighteen	eighty
9	nine	nineteen	ninety

Solid and dashed lines indicate numerically close and far pairs respectively.

1.2.2 The Spatial-Numerical Association of Response Codes (SNARC) Effect

The SNARC effect – the preferential rightward response with large numbers and leftward response with small numbers – may be seen as a numerical analogy to the Simon effect which refers to the increase in reaction time when the (task-irrelevant) location of the stimulus corresponds to the location of the response compared to when it does not. The effect, firstly identified and extensively investigated by Dehaene et al. (1993) in a series of experiments where subjects had to perform parity judgements, has been interpreted as a spatial congruency between the response side (left or right) and the relative position of numerical magnitudes on the mental number line. Regardless of parity status, large numbers were responded to faster with the right-hand key, whereas small numbers were faster with the left-hand key. This effect was equally significant with Arabic and verbal numerals (Dehaene et al., 1993; see also Fias, Brysbaert, Geypens, & d'Ydewalle, 1996), but very weak, if not absent, with two-digit numerals, regardless of notation, suggesting that single digits have privileged access to magnitude representations, perhaps because they are more frequent than larger numbers (with zero being the exception, Dehaene & Mehler, 1992).

More recently, Fias' (2001) study with verbal numerals has provided further insights into the SNARC effect. He used a parity judgement task and a phoneme monitoring task and observed the SNARC effect only in the former but not the latter. Fias (2001) concluded that the SNARC effect reflects semantic access in the parity task, whereas the phoneme monitoring task could "be performed through direct asemantic transcoding" and hence the absence of the SNARC effect.

Dehaene et al. (1993) reported that the SNARC effect did not vary with handedness but was related to the direction of writing. It was observed that Iranian subjects, as a group, showed little sign of the SNARC effect. Further analyses revealed a distinction between those subjects who had long been in France and learned a second (e.g., Western) language early in life and those who learned a second language late in life – the latter who were more familiar with a right-to-left writing system tended to associate large numbers with the left-hand side (i.e., weak or reversed SNARC effect), whereas the former tended to show the normal effect.

Dehaene et al. (1993, Experiment 4) carried out a consonant-vowel letter classification task, but the SNARC effect was absent, i.e., no preference for the right or left hand was found. The finding that the SNARC effect is specific to numbers, and not any stimuli obeying a fixed sequential order, such as letters, supports the notion that the effect depends on semantic (magnitude), and not ordinal, aspect of number representation.

Contrary to Dehaene et al.'s (1993, Experiment 4) finding that the SNARC effect is specific to numbers, Gevers, Reynvoet, and Fias (2003) observed the SNARC effect with non-numerical ordinal information. In an order-relevant task, subjects had to judge the position of months before and after July, and in an order-irrelevant task, they had to decide whether the presented month ended with the letter "R" or not (Gevers et al., 2003, Experiment 1). In another order-relevant task, subjects had to judge the position of the letters before and after "O", and in the order-irrelevant task, they had to classify consonants and vowels (Gevers et al., 2003, Experiment 2). The SNARC effect – the beginning of the sequence associated with the left side response and end of the sequence with the right side response – was observed in both order-relevant and -irrelevant tasks. The finding has demonstrated that the SNARC effect is not specific to numerical stimuli. It further suggests that the spatial organisation of numbers (as indicated by the SNARC effect) may reflect merely the ordinal rather than the semantic (magnitude) information.

In summary, the numerical distance effect can be used as a reliable indicator of semantic number processing, but the use of the SNARC effect to such a purpose appears somewhat unreliable. It is important to stress the difference between the tasks in which these effects have been observed – in number comparisons, numerical magnitude is the task-relevant dimension, whereas in parity judgements, numerical magnitude is the task-irrelevant dimension. If one assumes that the SNARC effect is a reliable indicator of semantic processing, then its presence during parity judgements can be taken as evidence for autonomous processing of numerical magnitude. However, it would be more appropriate to employ a paradigm where numerical distance effect can be examined while subjects make parity judgements (see Chapter 5). The key question here is whether access to numerical magnitude during parity judgements is inevitable since "numerical magnitude is more readily available than parity information" (Dehaene et al., 1993). Dehaene et al.'s (1993) memory retrieval account implies the direct retrieval of parity information bypassing numerical magnitude.

1.3 Experimental Paradigms Used in Studying Numerical Magnitude Processing

The current section discusses a few of the most widely used experimental paradigms in investigating numerical magnitude processing with regard to task relevance. As mentioned before, number comparison and parity judgement tasks have been used to study numerical magnitude processing. However, since parity depends on numerical magnitude (by definition, an even number is one which can be divided by 2 leaving no remainder, whereas an odd one refers to one which, when divided by 2, leaves 1 as a remainder), and that research has suggested that "numerical magnitude is more readily available than parity" (Dehaene et al., 1993), parity judgements cannot be reliably considered as a task probing numerical magnitude processing under task-irrelevant conditions. However, findings from parity judgement tasks can be used to compare those which tap specifically numerical magnitude processing under task-irrelevant conditions, such as number/ size comparison tasks and enumeration Stroop tasks.

1.3.1 Comparison Stroop Paradigm

Number comparison tasks have been modified to a Stroop paradigm, through varying the physical size of the numbers, thus allowing one to investigate numerical magnitude processing under both task-relevant and -irrelevant conditions. In a typical comparison Stroop study, two multidimensional numbers are compared along one of these dimensions: numerical magnitude and physical size. In such an experiment, subjects would be asked to judge which of the two digits is larger (or smaller) either in numerical magnitude or in physical size (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik & Tzelgov, 1982; Tzelgov, Meyer, & Henik, 1992; Yurko & Hinrichs, 1978). Trials may be congruent, where the numerically larger number is physically larger (e.g., 3 5); incongruent, where the numerically larger number is physically smaller (e.g., 3 5); and, in some experiments, neutral where the numbers are displayed in the same size (e.g., 3 5) for the numerical comparison task and where the same numbers are displayed in different sizes (e.g., 3 3) for the physical comparison task. The general findings are as follows. The Stroop effect manifests as interference (indicated by an increase in mean reaction time/ error rate across incongruent and neutral trials) and/ or facilitation (indicated by a reduction in mean reaction time/ error rate across neutral and congruent trials) (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik &

Tzelgov, 1982; Tzelgov et al., 1992). It is important to note that facilitation is virtually always substantially smaller than interference (see review, MacLeod, 1991).

Besner and Coltheart (1979) observed the Stroop effect with numbers (i.e., an increase in reaction times with incongruent trials compared to neutral and congruent ones). Henik and Tzelgov (1982) replicated Besner and Coltheart's (1979) findings with a numerical comparison task. In addition, they introduced a physical comparison task, where subjects had to select the physical larger one from a number pair, while ignoring the numerical magnitudes. They observed a weaker form of the Stroop effect in this task – only interference was observed, and only when this task was preceded by the numerical comparison task.

The introduction of physical comparisons was a highly important, though often underestimated, move in number research since this task specifically probes numerical magnitude processing under irrelevant conditions. It opened up a new gateway for studying numerical processing.

Henik and Tzelgov (1982) found that physical comparisons were made significantly faster than numerical comparisons. This finding was replicated by Tzelgov et al. (1992). Based on such a finding and the fact that, during physical comparisons, the slower (numerical) process influenced the faster (physical) one, Henik and Tzelgov (1982) concluded that "these processes are executed in parallel" but stated that "It is impossible, at this point, to claim that the numerical information is fully processed."

Further research has been carried out to investigate numerical information processing under task-irrelevant conditions. Henik and Tzelgov (1982, Experiment 2), observed the numerical distance effect, with incongruent trials, in numerical comparisons, and a reversed numerical distance effect – "the more distant (4 unit) pairs of digits were compared faster than the less distant (2 unit) pairs." Girelli et al. (2000) replicated Henik and Tzelgov's (1982) finding that a numerical distance effect was observed in a numerical comparison task, where numerical magnitude was the relevant dimension, and that a reversed numerical distance effect was observed in a physical comparison task, where numerical magnitude was the task-irrelevant dimension. Girelli et al. (2000) also observed that distant pairs exhibited a stronger Stroop effect than close pairs in numerical comparisons. The difference was more observable in physical comparisons, where only distant pairs led to the Stroop effect and Girelli (1998) suggested that "differentiating between close pairs, being more time-demanding, would not have been completed in time for interfering with the response". Similarly, in numerical comparisons, numerically distant pairs take less time to process than numerically close pairs (the classical distance effect), so by analogy, when numerical magnitude was the task-irrelevant dimension (in physical comparisons), numerically close pairs, being more time-demanding, would not exert as much influence on physical comparisons as numerically distant pairs, hence the reversed numerical distance effect.

At this point, it is important to note that the reversed numerical distance effect observed during physical comparisons in Henik and Tzelgov's (1982) and Girelli et al.'s (2000) studies has often been overlooked by researchers. This effect, in fact, highlights precisely that the processing of numerical magnitude was both refined (i.e., on a graded scale) and autonomous (since numerical magnitude was the task-irrelevant dimension). However, inconsistent findings (Tzelgov et al., 1992, Experiment 3; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002) have been reported, discrediting the validity of the effect. Experiments 2, 3a and 3b in the present thesis incorporated modifications into the existing comparison Stroop paradigm and attempts were made to replicate the reversed numerical distance effect with well-matched dimensions.

Tzelgov et al. (1992, Experiment 3) and Rubinsten et al. (2002) observed the numerical distance effect in a numerical comparison task only, but not a physical comparison task. Tzelgov et al. (1992) argued that the absence of the numerical distance implied that numerical information processing was less refined[7] in

[7]"Refined" information processing, in the present thesis, is reflected by a distance effect (in either the numerical or the physical dimension). This contrasts with the coarse large-small dichotomous classification, reflected merely by the Stroop effect.

physical comparisons where numerical magnitude was the irrelevant dimension – "although numerical information affects physical judgments, this influence reflects mainly the large versus small classification of numerical stimuli and not their exact numerical size". According to Tzelgov et al. (1992), digits 1-4 were classified as "small" and digits 6-9 were classified as "large".

In all the comparison Stroop experiments discussed above (except Rubinsten et al., 2002), factorial designs were employed, where nine numbers have been put into conflict with, at most, three sizes (usually a large size and a small size for constructing congruent and incongruent trials, and a medium size for neutral trials). This meant that the competing dimensions were not properly matched since there were many possible distances in the numerical dimension but two, at most, in the physical. This limited inferences that could be implicated on the amount of information available in the task-relevant and -irrelevant dimensions. Thus, physical comparisons only involved a large-small dichotomous classification rather than judgements on a graded scale. In fact, Rubinsten et al. (2002) used two physical distances (where the physical sizes of the digits differed either by 1 mm or 2 mm), but they did not test for the physical distance effect.

Moreover, the comparison Stroop tasks discussed above are very different: physical size comparison tasks can be solved perceptually, whereas numerical magnitude comparison tasks need to be solved conceptually. Given the difference in nature, these dimensions appear not very well-matched. In contrast, another numerical Stroop variant – the enumeration Stroop paradigm – which has been less widely employed, involves two dimensions, numerosity naming and reading Arabic numerals, both of which require conceptual processing.

1.3.2 Enumeration Stroop Paradigm

Windes (1968) first documented the interference effect in numerosity naming and reading Arabic digits; the numerical value represented by a digit could match or mismatch the number of items being displayed (i.e., the numerosity), for example, 2 2 versus 2 2 2 respectively. Neutral trials, where the items have no numerical value (e.g., X X X), may be included. Windes' (1968) finding was replicated by several studies (e.g., Flowers, Warner, & Polansky, 1979; Shor, 1971). Similarly, using a card-sorting task in which participants were required to order a group of cards according to the number of symbols printed on them, Morton (1969) observed interference with both Arabic and verbal numerals.

Although tasks like the one employed by Windes (1968, see also Bush, Frazier, Rauch, Seidman, Whalen, Jenike, Rosen, & Biederman, 1999; Bush, Whalen, Rosen, Jenike, Mclnerney, & Rauch, 1998; Parry, Scott, Palace, Smith, & Matthews, 2003; Pavese & Umiltà, 1998) are generally referred to as counting Stroop tasks, it is important to point out that in such a task, subjects do not necessarily count the number of items displayed; instead they subitize[8] smaller numerosities. Therefore, it is more appropriate to refer them as enumeration Stroop tasks since enumeration includes both counting and subitizing.

In a thorough study, Pavese and Umiltà (1998) examined the Stroop and the numerical distance effects in both subitizing and counting tasks. In the subitizing task, 5 digits (1 to 5) and 5 numerosities (1 to 5) were used, whereas in the counting task, the digits used were 5 to 9 and numerosities were also 5 to 9. Trials could be congruent, neutral, or incongruent (e.g., 3 3 3, Y Y Y, and 2 2 2 respectively). Subjects had to report vocally the numerosity of each display, while ignoring the identity of the stimuli.

Pavese and Umiltà (1998) observed that enumeration of small numerosities required significantly less time than that of larger ones. In the subitizing task, reaction time was observed to increase minimally but significantly with numerosity. However, in the counting task, reaction time was found to increase sharply and significantly with numerosity. These findings support the idea that

[8]The term "subitizing" was coined by Kaufman, Lord, Reese, and Volkmann (1949) to describe the rapid, confident, and accurate report of the numerosity of arrays of elements presented for short durations. They noted that this process, different from counting and estimating, was restricted to arrays with 6 or fewer elements.

there are two enumeration mechanisms: subitizing which involves simple pattern-matching based on recognition and counting which relies on the existence of counting knowledge, represented in the form of an ordered sequence of number facts (Peterson & Simon, 2000).

The Stroop effect, manifested as interference (an increase in reaction time across neutral and incongruent trials) and facilitation (a reduction in reaction time across neutral and congruent trials), was observed in the subitizing task. However, in the counting task, neutral trials were responded to slower than incongruent trials, making the former an unreliable baseline, hence the Stroop effect could not be established.

With regard to the distance effect, Pavese and Umiltà (1998) found that reaction times for incongruent close trials were significantly slower than those for incongruent distant trials (e.g., 3 3 3 3 and 1 1 1 1 respectively) in both subitizing and counting tasks, indicating that close trials interfered more strongly than distant trials.

In the experiment, Pavese and Umiltà (1998) employed a design where the stimulus array was displayed until a response was given. Such a procedure might have allowed participants to process task-irrelevant digit identity at a later stage, i.e., after the enumeration process has been completed. So, Pavese and Umiltà (1999) carried out another experiment, employing a slightly different design so that the stimulus array was briefly presented and masked. Even with such a short exposure time, the numerical distance effect was replicated – incongruent close digits were enumerated more slowly than incongruent distant digits. Such a finding suggests that the numerical distance effect depends on a rapid and automatic (rather than late) activation of magnitude information from the task-irrelevant digit identity.

Although the enumeration paradigm has the advantage of employing two dimensions of similar nature (namely, numerical magnitude and numerosity) both of which involve conceptual processing, there are major drawbacks concerning the paradigm.

Firstly, the paradigm suffers the same criticism by Zysset, Müller, Lohmann, and von Cramon, (2001) as the traditional colour-word Stroop paradigm, "generating the verbal response to match a stimulus is interfered by the second dimension of the stimulus (or the dimension of a second stimulus). Response preparation and the interference process itself are confounded by this." In the colour-word Stroop paradigm, for example, generating the verbal response "red" to match one dimension, namely colour, to the stimulus "GREEN" is slowed by the second dimension of the stimulus, namely the identity of the word. Likewise in the enumeration Stroop paradigm, generating the verbal response "four" to match one dimension, namely numerosity, to the stimulus set "3 3 3 3" is slowed by the second dimension of the set, namely the identity of the digits.

Secondly, response preparation and response suppression cannot be well-controlled for in a paradigm which makes use of vocal responses. Vocal response preparation is different for each different digit, and more importantly, it is different for each suppressed digit. For example, stimulus sets, "2 2 2" and "5 5 5" both require a correct answer of "three", however, one needs to suppress the preparation for "two" in the former, but "five" in the latter. A better methodology would be to introduce a matching process in which subjects would be able to respond by key presses. This will be further discussed in Chapter 6.

In summary, both the comparison Stroop and the enumeration Stroop paradigms are useful tools in studying the processing of numerical magnitude information under task-irrelevant condition, but modifications are needed to improve these paradigms.

1.3.3 Parity Judgement Paradigm

Parity, besides numerical magnitude, is another important property of numbers. However, parity information processing has not been as extensively researched as magnitude. The parity judgement paradigm has often taken the simple form of determining whether a given number is even or odd. A common finding is that odd numbers are responded to slower than even ones (e.g., Dehaene et al., 1993; Hines, 1990). There is also consensus towards the finding that numerical magnitudes are more available than parity information. For details of research on parity judgements, see Section 5.1. A more elaborate parity judgement paradigm has been devised to investigate the processing of numerical magnitudes when these were presented as task-irrelevant (see Chapter 5).

1.4 Models of Numerical Representation and Processing

In recent years, there has been a growing interest in the cognitive mechanisms underlying numerical processing calculation abilities. Several models have been proposed (e.g., Campbell & Clark, 1988; Cipolotti & Butterworth, 1995; Dehaene, 1992; McCloskey, 1992; McCloskey, Caramazza & Basili, 1985), some of which attempt to provide a general architecture of numerical processing (e.g., Campbell & Clark, 1992; Dehaene, 1992; McCloskey et al., 1985), others contribute to specific aspects of it (e.g., Cipolotti, 1995; Cipolotti & Butterworth, 1995; Cohen et al., 1994; Deloche & Seron, 1982a, b, 1987; Noël & Seron, 1993; 1997; Zorzi & Butterworth, 1999). The evidence in support of the different theories comes from experimental investigations with normal subjects, neuropsychological single-case studies, and computational simulations.

At present, there is a general consensus towards basic observations grounded on solid and replicable empirical findings, e.g., the numerical processing system is modular and there are distinct modules responsible for basic number processing and calculation. However, some fundamental issues are still controversial, in particular, the format in which numbers are mentally represented and the extent to which a semantic representation is central to numerical processing. The current thesis focuses on semantic processing of numbers, hence it is important to consider how numbers are represented. However, numerical processing via asemantic routes and issues relating to calculation will not be discussed in detail (for discussion, see Cipolotti & Butterworth, 1995).

In the following sections, the current models of numerical processing will be described and emphasis will be given to their specific assumptions related to numerical magnitude representation and processing.

1.4.1 McCloskey's Semantic Model and the Incorporation of Asemantic Transcoding Pathways

An influential model was first proposed by McCloskey et al. (1985) and subsequently refined in several works (Macaruso, McCloskey, & Aliminosa, 1993; McCloskey, 1992; McCloskey, Aliminosa, & Sokol, 1991; Sokol, Goodman-Shulman, & McCloskey, 1989; Sokol & McCloskey, 1991; Sokol, McCloskey, Cohen, & Aliminosa, 1991). The model is characterised by a modular architecture and by the pivotal role of the abstract internal system which is semantic in nature. It posits a first basic distinction between the number processing system and the calculation system (see Figure 1.1).

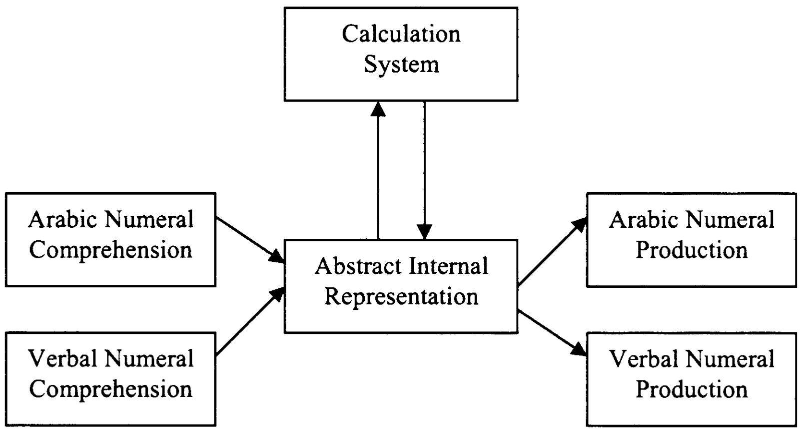


Figure 1.1 McCloskey's semantic model

The abstract internal representations, according to McCloskey's model, are modality-independent and semantic in nature, and are assumed to specify the basic quantities of numbers and their associated power of ten, for example, the internal abstract representation of 5493 would be as follows: {5}10EXP3, {4}10EXP2, {9}10EXP1, {3}10EXP0. The digits in the brackets (e.g., 5 in {5}) stands for the semantic representation of the number and 10EXPn specifies the power (n) of ten associated with each quantity (e.g., 10EXP3 refers to 10 to third power).

The internal representations constitute the input to and the output from the calculation module. McCloskey's model assumes that all numerical inputs (e.g., 6, six, or /siks/) are initially translated, via notation-specific comprehension modules, into amodal abstract representations of numbers. Conversely, number production involves a translation from the abstract internal representations to the desired output notation either verbal (spoken or written) or Arabic, via notation-specific production modules.

Single-case studies of cerebral-lesioned patients presenting selective difficulties in different aspects of number processing provide support for the functional distinctions depicted in McCloskey's model (see McCloskey, 1992 for a review). For example, patients showing a dissociation between intact comprehension and impaired production processes have been reported (Benson & Denckla, 1969; McCloskey et al., 1986; Singer & Low, 1933), as well as patients with a selective deficit in the comprehension of verbal numerals and intact comprehension of Arabic digits (McCloskey & Caramazza, 1987; McCloskey et al., 1985, 1986).

In summary, McCloskey's model postulates that performance of any transcoding task is accomplished via obligatory activation of the internal semantic representations. The same assumption applies to calculation and other numerical tasks, such as number comparisons, and the model predicts that for a given task, e.g., the comparison Stroop task, format-related differences in the performance would result only from differences in the encoding stage, thus the Stroop effect should manifest with both Arabic digits and written verbal numerals and this is indeed the case (e.g., Foltz et al., 1984); the difference between these stimuli only lies in the reaction times – Arabic digits were responded to faster than written verbal numerals.

However, the assumption that the number production mechanism could only be accessed through abstract semantic internal representations has been challenged by several neuropsychological case studies (Cipolotti, 1993, 1995; Cipolotti &

Butterworth, 1995; Cohen et al., 1994; Noël & Seron, 1995). In particular, Cipolotti (1995) reported a patient SF who showed a selective deficit in reading aloud Arabic numerals in the absence of comprehension and production problems. Also, Cipolotti and Butterworth (1995) reported a patient SAM who presented a dissociation between impaired verbal and Arabic numeral production in transcoding tasks and preserved spoken and Arabic numerals production in calculation tasks. In both of these cases, the observed difficulties to produce a particular number code were task-specific. Cipolotti and Butterworth (1995) proposed that numeral output systems can be accessed not only via abstract semantic representations, but also through asemantic routes that bypass abstract internal representations. Cipolotti and Butterworth's (1995) model will be further discussed in Section 1.4.5.

1.4.2 The Incorporation of Asemantic Routes in Models of Numerical Processing

Other models have also explicitly rejected McCloskey's hypothesis of a pivotal abstract number representation. In particular, Deloche and Seron (Deloche & Seron, 1982a, b, 1987; Seron & Deloche, 1983, 1984) have repeatedly argued for the possibility of asemantic transcoding, that is, direct translation between Arabic and verbal notations without going through an intermediate semantic representation. They developed a detailed model of the processes underlying transcoding of written verbal numerals to Arabic numerals and vice versa.

More recently, Noël and Seron (1993) incorporated Deloche and Seron's asemantic transcoding algorithms into their preferred entry code hypothesis. According to these authors, the access to number semantic representations may be accomplished either from the verbal or the Arabic code, according to the individual's idiosyncratic preference in maintaining the information in his/ her working memory in an auditory or a visual code. Thus, before any numerical task can be performed, transcoding is carried out from the input format and to the preferred format and this is done through asemantic mechanisms. The hypothesis of a verbal pivotal representation for numbers is also supported by the anecdotal observation that bilinguals prefer to perform calculations in the language in which they acquired and practised arithmetic facts (Kolers, 1968). Unfortunately, this hypothesis lacks solid, empirical evidence.

1.4.3 Campell and Clark's Encoding Complex View

The most extreme non-modular model of numerical processing has been processed by Campbell and Clark (Campbell, 1994; Campbell & Clark, 1988, 1992; Clark & Campbell, 1991). Their interactive architecture does not assume a central abstract representation. Instead, it postulates that numbers evoke "an integrated network of format-specific number codes and processes that collectively mediate number comprehension, calculation, and production". The model assumes both verbal codes (e.g., auditory, orthographic, and articulatory) and non-verbal representations (e.g., visual and analogue) and they are interconnected in an associative network so that any code/ representation could potentially activate or inhibit other ones at any give time.

The model postulates that "different number formats can differentially activate internal number representations and associations" (Campbell & Clark, 1992), hence predicting format effects in number processing tasks. However, as mentioned before, the Stroop effect has been observed with both Arabic digits and written verbal numerals (e.g., Foltz et al., 1984). Also, Dehaene and Akhavein (1995) reported the numerical distance effect in a same-different judgement task with both Arabic digits and written verbal numerals. These findings are inconsistent with the prediction of format effects in number processing.

It must be noted that, according to Campbell (1992), the absence of a central amodal representation does not entail the rejection of a numerical semantic representation. Within this framework, magnitude is not a unitary concept, "rather the representation of magnitude corresponds to a set of specific learned relations and processes (e.g., labelling of perceptual groups or intensities, use of counting and other basic relations to present changes in quantity)" (Campbell, 1992). Moreover, semantic representations are assumed to be modality-specific, for example, visuo-spatial (in the form of a number line) or verbal (in the form of well-learned counting sequence). In this respect, it follows to assume multiple transcoding routes. Thus, in principle, although not explicitly postulated, the encoding complex view postulates both semantic and asemantic pathways. Unfortunately, the flexibility arising from the unspecified pathways and mechanisms renders Campbell and Clark's model largely untestable in its full generality (Dehaene et al., 1993; McCloskey, Macaruso, & Whetstone, 1992).

1.4.4 Dehaene's Triple-Code Model

Consistent with the modularity assumption postulated by McCloskey and colleagues, Dehaene (1992) proposed the triple-code model for number processing. The framework consists of three types of mental numerical representations, namely, visual-Arabic number form in which numbers are represented as string of digits, auditory-verbal number form in which numbers are represented as syntactically organised sequence of words, and analogue magnitude code in which magnitude information is represented as distribution of activation on an oriented number line (see also Restle, 1970).

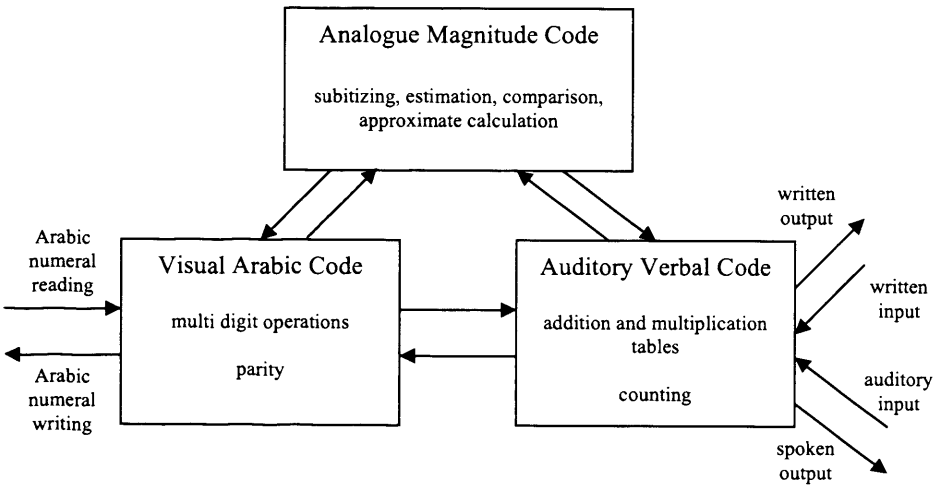


Figure 1.2 Dehaene's triple-code model for number processing (after Dehaene, 1992)

The model assumes that each representation mediates specific numerical activities. For example, the visual-Arabic representation is postulated to support multi-digit operations as well as to have exclusive access to parity information (e.g., Dehaene et al., 1993), the auditory-verbal code to mediate verbal input and output, to support counting, and to provide the medium of representation for memorised addition and multiplication facts (Dehaene & Cohen, 1997), the analogue representation to support tasks which require quantity information, e.g., number comparison and approximate calculation (e.g., Cohen & Dehaene, 1991; Dehaene, 1989; Dehaene et al., 1990).

According to the triple-code model, for any given task, the very first stage is to translate the input stimulus into the appropriate representation. Given that the representations are linked to one another, the triple-code model posits both semantic and asemantic transcoding routes. As depicted in Figure 1.2, Arabic and verbal codes are connected by direct asemantic pathways, as well as through semantic analogue magnitude representation.

The support for the assumption of direct and indirect transcoding routes comes mainly from neuropsychological data (e.g., Cohen & Dehaene, 1991, 1995; Cohen et al., 1994). For example, Cohen et al. (1994) reported a deep dyslexic patient whose difficulties in reading numbers were constrained to unfamiliar numerals while "meaningful" numbers, such as famous dates, were better memorised. This lends support to the distinction between direct (asemantic) and indirect (semantic) mechanisms of numerical processing.

More recently, Dehaene and Cohen (1997) reported two acalculic patients who, despite having preserved transcoding abilities, showed very different patterns of impaired numerical knowledge. Patient BOO suffered from a localized lesion in the left basal ganglia, a circuit which has been postulated to be partially involved in memory for rote verbal material such as arithmetic facts (Dehaene & Cohen, 1995). In good agreement with Dehaene's triple-code model, BOO showed most difficulties with multiplication[9] and division and was severely impaired at rote verbal knowledge (e.g., reciting the alphabet and the musical scale), but had no difficulties in a series of tasks based on quantitative numerical knowledge (e.g., number comparison, approximate calculation). In contrast, patient MAR who suffered a localized inferior parietal lesion, a lesion which according to Dehaene's triple-code model should spare the retrieval of rote verbal knowledge, showed a "preservation of rote arithmetic tables" but was impaired at number comparison and approximate calculation. These findings seem in line with Dehaene's triple-code model. However, upon detailed inspection, the presumed double dissociations observed in Dehaene and Cohen's (1997) patients were less clear-cut. For example, despite the claim of a "preservation of rote arithmetic tables", patient MAR's accuracy varied from 25% to 73% across the four arithmetic operations. Despite the weak support for a distinction between semantic and asemantic routes provided by the unclear double dissociation, further patients have been reported whose behaviours pose difficulties for a single-route semantic model (see Section 1.4.5 for details).

More central to the current discussion is the assumption of an analogue magnitude representation, which is pictured as an oriented and compressed number line, originally proposed to account for the distance effect in number comparison (or more precisely, numerical magnitude comparison), thus the time to compare a number pair is explained by a logarithmic function of the numerical distance between them (Aiken & Williams, 1968; Moyer & Landauer, 1967; Restle, 1970). Dehaene's (1992) model rests on Moyer and Landauer's (1967) assumption that "the decision process... is one in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line" (see also Dehaene, 1992; Dehaene & Cohen, 1995; McCloskey, 1992).

Support for the analogue magnitude representation comes mainly from experimental data with normal subjects. In a 2-digit comparison task, even in cases where the comparison of the decades was sufficient for the judgement, the

[9]In normal adults, multiplication is solved largely by memory retrieval (Ashcraft, 1992) and division almost certainly requires a search through multiplication tables.

units had an influence on the reaction times (Dehaene et al., 1990). Such a finding supports Dehaene's model that the decades and units are first combined in a unique analogue magnitude representation rather than processed serially at the digit level.

Further support for the analogue magnitude representation comes from a series of experiments on parity judgements (Dehaene et al., 1993). It was observed that "Large numbers preferentially elicited a rightward response, and small numbers a leftward response" and this effect has been named the Spatial-Numerical Association of Response Codes (SNARC) effect. This effect was found to be unrelated to handedness or hemispheric dominance, but was linked to the direction of writing, "as it faded or even reversed" in right-to-left writing Iranian subjects (Dehaene et al., 1993). These findings support the notion that mental representation of numbers is spatially organized on a mental number line.

However, Vuilleumier et al.'s (2004) research with neglect patients has cast some doubts onto Dehaene's analogue magnitude representation of numbers. Although in two number comparison tasks, lateralized disturbance in the mental representation of numbers was observed in neglect patients – "When asked to judge whether a single number shown at fixation was smaller or larger than "5", patients with neglect were selectively slower to respond to "4", but when asked to compare numbers to "7" they were selectively slower to respond to "6".", supporting that idea that mental representation of numbers is spatially organized, these patients did not show a selective loss in a number Stroop task (where patients had to judge whether the physical size of the presented number was physically small or large) with leftward numbers (below 5). The intact number Stroop effect in physical size comparisons suggest that numerical magnitudes were processed by the patients under task-irrelevant conditions. These findings suggest a distinction between cardinal and ordinal representations of numbers – while the neglect patients had intact cardinal representation (evoked when processing numerical magnitudes), they showed lateralized disturbance with left side/ numerically smaller numbers (i.e., the ordinal representation).

The analogue magnitude representation view has been criticized by Butterworth and colleagues (e.g., Delazer & Butterworth, 1997; Zorzi & Butterworth, 1999). "Analogue representations, however, fail to capture our intuitive notion of whole numbers, and whole-number arithmetic. Perhaps because of our early experience with counting, we intuitively think of whole numbers as meaning not approximate analogue magnitudes, but discrete numerosities" (Zorzi & Butterworth, 1999). The idea that a whole number denotes the numerosity (or cardinality) of a set will be further discussed in Section 1.4.6.

Despite the criticism, Dehaene's triple-code model may be considered successful in its attempt to provide a structured and testable functional architecture of numerical processing. It has the merit of explicitly referring to numerical abilities which had previously been neglected (e.g., approximation, counting).

1.4.5 Cipolotti and Butterworth's Multiroute Model

Cipolotti and colleagues (Cipolotti, 1993, 1995; Cipolotti & Butterworth, 1995; Cipolotti, Butterworth, & Warrington, 1995) proposed a modified version of McCloskey's model in order to account for patients' performance that would have been problematic to interpret within a single-semantic route model. In particular, patient SF showed a selective deficit in reading aloud Arabic numerals in the absence of comprehension and production problems (Cipolotti, 1995), and patient SAM presented a dissociation between impaired verbal and Arabic numeral production in transcoding tasks and preserved spoken and Arabic numeral production in calculation tasks (Cipolotti & Butterworth, 1995). In both cases, the observed difficulties in producing a particular number code were task-specific and pose difficulties for a single-route semantic model. Hence, Cipolotti and Butterworth (1995) proposed a multiroute model to reconcile these findings, critically, numeral output systems can not only be accessed via abstract semantic representations, but also through asemantic routes which bypass the abstract internal representation (see Figure 1.3).

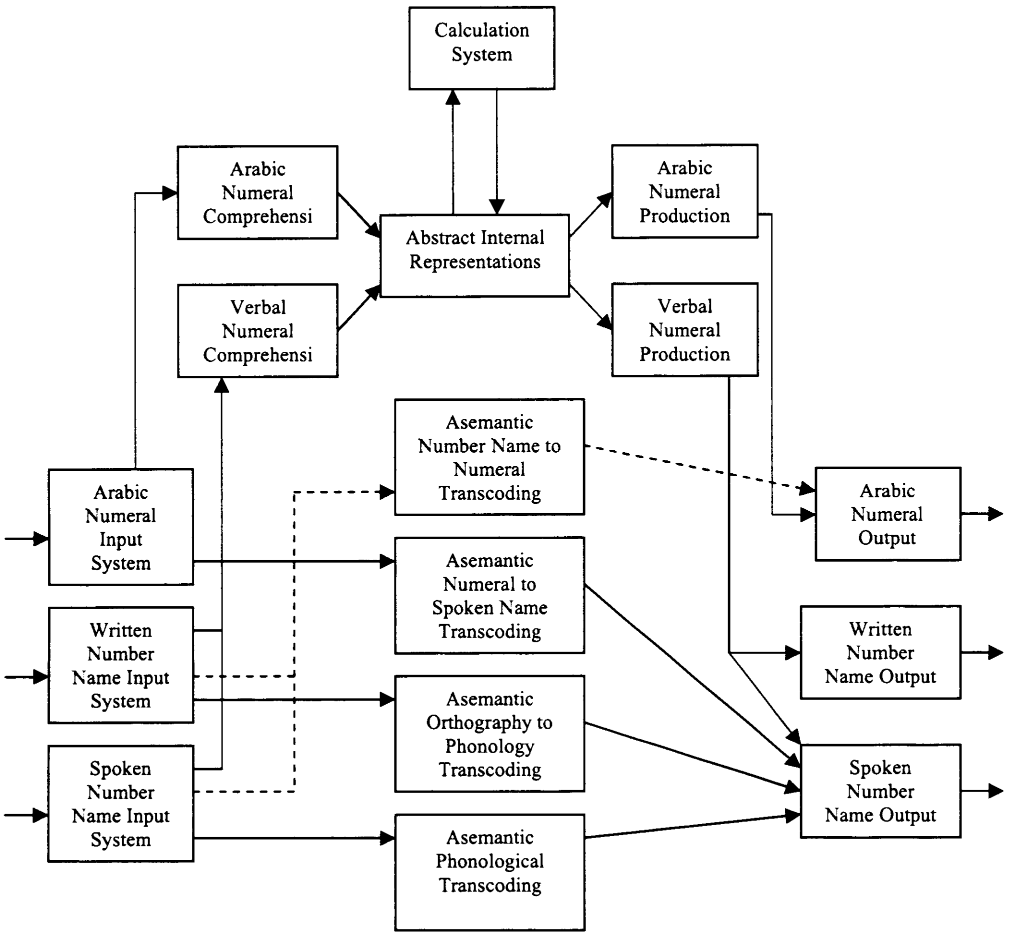


Figure 1.3 Cipolotti and Butterworth's (1995) multiroute model

1.4.6 Zorzi and Butterworth's Linear Representation Model

As mentioned before, Zorzi and Butterworth (1999) criticized Dehaene's analogue representation, based on a compressive mental number line, for failing "to capture our intuitive notion of whole numbers, and whole-number arithmetic". Instead, they proposed a linear representation of number magnitudes. According to these authors, "number representations are ordered by numerosity: smaller numbers denote proper subsets of the sets denoted by bigger numbers." This hypothesis entails that larger numbers are more similar to each other than smaller numbers, without assuming a logarithmic compression, since large numbers share more active nodes. For example, the pairs (2 5) and (5 8) both consist of numbers with a numerical distance of 3, but the former pair of numbers share only 2 nodes while the latter numbers share 5 (see Figure 1.4). In other words, the size effect arises as a result of the increasing number of shared nodes with larger numerosity representations.

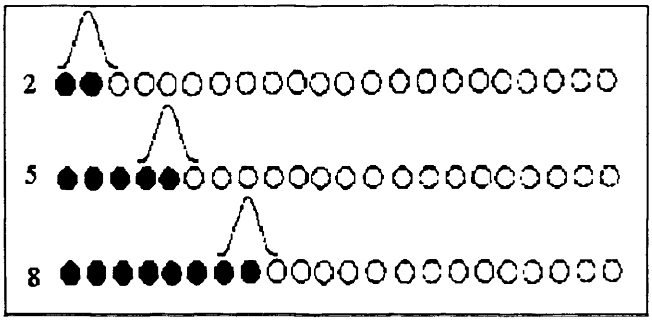


Figure 1.4 Zorzi and Butterworth's (1999) numerosity representations

Using computer simulation, Zorzi and Butterworth (1999) readily demonstrated, in a number comparison task, the size effect (i.e., for a given distance, comparison is faster for smaller numbers (e.g., 2 and 4) than for larger numbers (e.g., 7 and 9), e.g., Moyer & Landauer, 1967) and the numerical distance effect (i.e., reaction time increases as the difference between the two numbers becomes smaller, e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977) which resembled the performance of human subjects. The numerical distance effect was explained by the stronger competition between two close (hence more similar) numbers than the weaker competition between two distant (hence less similar) numbers; the former would produce a longer reaction time. Within this framework, numerals are mapped linearly on to magnitude representations, "and the compressive effect on the comparison times emerges by virtue of the non-linear interactions that are intrinsic to the decision process itself" (Zorzi & Butterworth, 1999).

Further support for Zorzi and Butterworth's (1999) model comes from the simulation data of Zorzi, Stoianov, Priftis, and Umiltà (2003) who reported a symmetrical distance effect in a number naming experiment, mimicking human performance (Reynvoet, Brysbaert, & Fias, 2002) – the reaction time to name the target number (e.g., 5) does not differ significantly if it has been primed by a smaller or larger number with the same numerical distance to the target (e.g., 3 or 7) (Reynvoet et al., 2002).

Zorzi and Butterworth's (1999) model differs from an earlier neural network model proposed by McCloskey and Lindemann (1992). According to the latter, numbers are encoded over an ordered sequence of input nodes, where each node stands for a particular number. More importantly, McCloskey and Lindemann's (1992) model assumes that the two immediate neighbours of a given number are activated as well. For example, 5 is represented as the activation of the node labelled "5" plus (lesser) activation of "4" and "6". Such representations are generally known as the "barcode" magnitude representations (see also Anderson, Spoehr, & Bennett, 1994; Viscuso, Anderson, & Spoehr, 1989). Although McCloskey and Lindemann's (1992) model provides some ordering of numbers (e.g., 5 has 4 and 6 as neighbours), numbers of larger numerical distance would activate orthogonal representations (e.g., 8 would activate nodes 7, 8, and 9, while 4 would activate nodes 3, 4, and 5) with no overlapping nodes. Zorzi and Butterworth (1999) questioned the feasibility of McCloskey and Lindemann's (1992) model in capturing the distance effect and criticised their lack of attempt to model number comparison.

According to Zorzi and Butterworth's (1999) linear representation model, comparison is between discrete numerosities (representations evoked by numbers), whereas Dehaene's triple-code model assumes that comparison is performed between analogue representations of numbers.

Evidence for a distinction between continuous quantity and discrete numerosity representations comes from a study by Castelli et al. (2006) who devised the Discrete Analogue Response Task (DART) which involved subjects viewing hues which changed abruptly (discrete, numerous stimuli) or smoothly (analogue/ continuous, non-numerous stimuli) in space or in time, and had to decide whether they saw more green or more blue. Enhanced activity was observed along the intra-parietal sulcus when viewing discrete stimuli compared to viewing continuous stimuli. The authors suggested that estimation of numerosity is a distinct process from perceiving continuous quantity and that an intra-parietal network connects segmentation of objects to estimation of their numerosity.

1.4.7 Walsh's A Theory of Magnitude (ATOM)

Following Gallistel and Gelman (2000) who argued that "countable and uncountable quantity (numerosity and amount, duration, etc.) should be represented with the same kind of symbols (mental magnitudes)," Walsh (2003) proposed A Theory of Magnitude (ATOM), linking time, space, and quantity. The author has proposed that there is a common magnitude system for processing information in the above dimensions and that the inferior parietal cortex is the locus of the system.

Evidence for a common magnitude system comes from both behavioural and neuroimaging studies, as well as neuropsychological cases, and may be divided into three sections where information processing shows an overlap between (1) space and quantity, (2) time and quantity, and (3) time and space.

Walsh (2003) argued that the distance effect observed in number comparison tasks (e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977) may be viewed as evidence for a common magnitude system for processing spatial and numerical information. According to the mental number line proposal (Dehaene, 1992; Dehaene & Changeux, 1993; Dehaene et al., 1990; see also McCloskey, 1992; Restle, 1970), subjects perform comparisons between the locations of the numbers on the mental number line.

Further evidence for a common magnitude system for processing spatial and numerical information comes from an unexpected feature of neglect uncovered by Zorzi et al. (2002). The authors reported that the mental bisection between two numbers was systematically shifted to the right (i.e., towards larger numbers) in neglect patients, paralleling their bias when asked to mark the centre of a physical line (Marshall & Halligan, 1989). The finding is consistent with the suggestion that there is an anatomical proximity or overlap between the brain areas presenting number and space.

The behavioural evidence for a common system for processing temporal information and quantity rests primarily on dual-task experiments. For example, Brown (1997), using three dual-task conditions, found that all secondary tasks (rotor tracking, visual detection, and mental arithmetic) disrupted performance on the temporal task, but only mental arithmetic was impaired by the temporal task. Walsh (2003) argued that the latter two tasks draw upon common magnitude mechanisms, whereas the other two tasks make predominantly visual demands on the subjects.

Further evidence for a common system for processing temporal information and quantity comes from neuropsychological cases. For example, Basso, Nichelli, Frassinetti, and di Pellegrino (1996) reported that a patient with left spatial neglect tested on short-duration estimation tasks (300 vs. 700 ms) consistently overestimated durations of stimuli presented in the neglected space and underestimated durations of stimuli presented in the unimpaired field.

The behavioural evidence for a common system between time and space comes from De Long's (1981) study, where subjects had to carry out tasks in environments built to 1/6, 1/12, or 1/24 of actual size and to stop when thirty minutes have passed. The ratio of time passed to time estimated scaled according to environmental scale (see also Mitchell & Davies, 1987).

There are also neuroimaging data to support a common magnitude system for time, space, and quantity; the parietal cortex is activated in many studies that investigated aspects of temporal, spatial, and number processes (e.g., Chochn, Cohen, van de Moortele, & Dehaene, 1999; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Piazza, Mechelli, Butterworth, & Price, 2002; Rao, Mayer, & Harrington, 2001; Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002).

Furthermore, studies using transcranial magnetic stimulation (TMS) have shown that parietal cortex stimulation in human subjects can cause deficits in spatial tasks (e.g., Ashbridge, Walsh, & Cowey, 1997; Bjoertomt, Cowey, & Walsh, 2002; Rushworth, Ellison, & Walsh, 2001), in number comparison (e.g., Göbel, Walsh, & Rushworth, 2001b), and in time discrimination (e.g., Walsh & Pasual-Leone, 2003).

Note that similar to Dehaene's analogue magnitude representation, Walsh's (2003) ATOM does not discriminate between discrete magnitudes and continuous quantities.

1.4.8 Verguts, Fias, & Steven's Model of Exact Small-Number Representation

To date, four main types of number representation have been proposed (Dehaene, 1992; Gallistel & Gelman, 1992; Verguts, Fias, & Stevens, 2005; Zorzi & Butterworth, 1999). The assumptions behind the different types of numerical representation are, however, different. Nevertheless, there is a general consensus that mental number representations are seen as organized along a mental number line – that is the numbers are represented with overlapping distributions of activation. This number line assumption can explain the numerical distance effect – comparison is slower for numbers that are numerically closer than those that are numerically more distant), since numbers close to each other (e.g., 1 and 2) have more distributional overlap than do numbers that are more distant (e.g., 1 and 4).

In contrast, each of the different types of number representation provides its own explanation to the size effect – for a given distance, comparison is faster for smaller numbers than for larger numbers. Firstly, Dehaene's (1992) analogue representation is based on a compressive mental number line implying that the distances between numerical representations on the number line are smaller for larger numbers. Such a compressed scaling entails that small numbers are represented as being further apart from each other than larger numbers for a fixed numerical distance, and therefore small numbers are easier to disambiguate from larger ones. Secondly, according to Zorzi and Butterworth (1999), whole numbers denote discrete numerosities by which number representations are ordered: smaller numbers denote proper subsets of the sets denoted by bigger numbers (i.e., larger numbers activate more nodes than smaller ones). Thus, given a fixed numerical distance, two larger numbers will share more active nodes than two smaller ones (e.g., 5 and 8 share 5 nodes, whereas 2 and 5 share only 2 nodes), thus the former are more difficult to disambiguate. Thirdly, according to Gallistel and Gelman (1992), the variability of number representations increases as magnitude increases. Two possible interpretations follow this view: 1) units further away are activated with smaller probability, or 2) units further away are activated less strongly. Either way, two larger numbers are more difficult to disambiguate due to the larger distributional overlap of activation than two smaller ones for a fixed numerical distance.

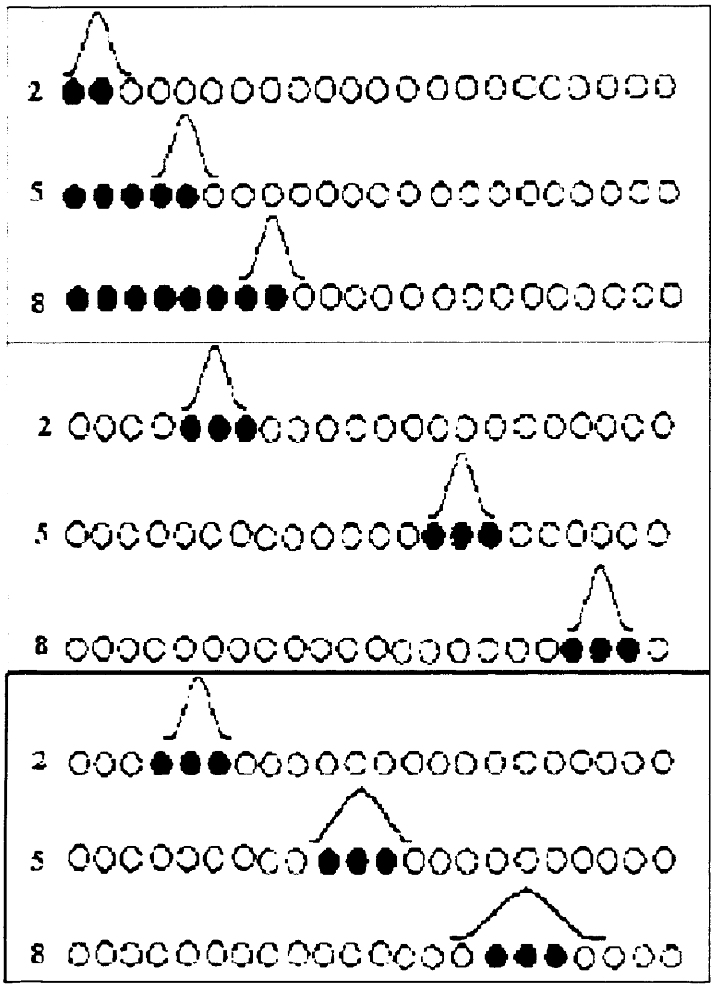


Figure 1.5 Three types of number representation: Zorzi and Butterworth's numerosity representation (top), Dehaene's compressive number line representation (middle), and Gallistel and Gelman's increasing variability representation (bottom)

Even though models based on any of the above number representations can explain the numerical distance effect and the size effect during number comparisons, Verguts et al. (2005) pointed out none of these is able to account for the symmetric priming effect in number naming and parity judgement – the reaction time to name the target number (e.g., 5) does not differ significantly if it has been primed by a smaller or larger number with the same numerical distance to the target (e.g., 3 or 7) (Reynvoet & Brysbaert, 1999, 2004; Reynvoet et al., 2002). Contrary to Verguts et al.'s (2005) claim, simulation data (Zorzi et al., 2003) based on Zorzi & Butterworth's (1999) model of discrete numerosity representations mimicked human performance in a number naming experiment – the distance effect was symmetrical (see also Section 1.4.6).

Verguts et al. (2005) argued mental number representations should be linear (i.e., non-compressive), should activate equal number of units (i.e., larger numbers should not activate more nodes than smaller numbers), and should have constant variability. Hence, they proposed a model of exact small-number representation which "preserves distances of the integer number line", e.g., the distance between the representations of 7 and 8 is the same as that between 1 and 2. This is consistent with Zorzi and Butterworth's (1999) "intuitive notion of whole numbers, and whole-number arithmetic". Evidence for Verguts et al.'s (2005) model comes from their simulation data which mimicked that of human performance: the numerical distance effect and the size effect in number comparisons (e.g., Moyer & Landauer, 1967) but symmetric priming effect in number naming and parity judgement (Reynvoet & Brysbaert, 1999, 2004; Reynvoet et al., 2002). The authors explained the data in terms of the differential frequencies of numbers in the training regime which resulted in nonlinearities in the connection weights between the number line and the comparison field, but not in the connection weights between the number line and the naming and parity output fields.

Verguts et al.'s (2005) model used a training regime in which small numbers were more frequent than large numbers and pointed out that the resultant number line would be one in which small numbers were represented exactly but representations would become fuzzier for larger numbers. However, Carey (2001) pointed out that it would make little sense to use fuzzy number representations to compare two large numbers such as 4,653 and 4,655. Moreover, neuroscientific evidence have suggested that brain circuits involved in the processing of small numbers might (partly) differ from those involved in processing large numbers (e.g., Göbel, Rushworth, & Walsh, 2001a; Göbel et al., 2001b; Whalen & Morelli, 2002). Verguts et al.'s (2005) therefore proposed that there might be two numerical systems: one which is used for exact small and approximate large numbers, and the other which represents multidigit numbers. The latter system decomposes multidigit numbers in its base-10 representation of different number lines, so that one number line represents the unit value, one number line represents the decade value, and so on. Recent data have provided support for such a decomposition for multidigit numbers (Nürk, Weger, & Willmes, 2001; Ratinckx, Brysbaert, & Fias, In Press).

1.4.9 Summary

The theoretical accounts of numerical representation currently reach the general consensus that the numerical processing system is modular and there are distinct modules responsible for basic number processing and calculation, that it consists of both semantic and asemantic pathways, the usage of which is task-dependent, and that there is an internal representation of numerical magnitude upon which processes such as number comparisons take place. However, the nature of such a representation is still undetermined. Different proposals have been made. Recent research suggests that numbers are linearly represented mentally (Zorzi & Butterworth, 1999; Verguts et al., 2005), challenging the previously popular concept of a compressive mental number line (Dehaene, 1992).

1.5 Theoretical Accounts of Numerical Magnitude Processing

The present thesis concerns the processing of numerical magnitude. As stated before, one of the most common paradigms used to study such a process is the number comparison task. By varying also the physical size of the numbers (hence the comparison Stroop paradigm), numerical magnitude can be manipulated as the task-irrelevant dimension. Thus, numerical magnitude processing can be investigated by studying under both task-relevant and task-irrelevant conditions, using numerical and physical comparison tasks respectively. This section discusses the theoretical accounts of number comparison.

1.5.1 Foltz et al.'s (1984) Relative Speed of Processing Account

Contrary to the previous finding that the Stroop effect only occurred with Arabic digits but not written verbal numerals (Besner & Coltheart, 1979), Foltz et al. (1984) observed the Stroop effect (manifested as interference and facilitation) with both types of stimulus using a repeated-set design (in which every number was paired with all other numbers). Foltz et al. (1984) attributed the absence of the Stroop effect with written verbal numerals in Besner and Coltheart's (1979) study to their use of a fixed-pairs design (in which each number was not paired with all other numbers, and number pairs were presented repeatedly); employing a limited number of stimulus pairs might have allowed responses to be made upon memory of previous trials rather than upon comparison per se.

Foltz et al. (1984) put forward a relative speed of processing explanation for their results, assuming a race between numerical magnitude and physical size comparisons (which involved similar processing stages but equivalent stages had different latencies for different stimulus types) and attributing the interference to the fact that the irrelevant dimension won the race. Their explanation was based on the following assumptions about the processing of both the numerical magnitude and physical size.

It was assumed that physical comparisons were relatively fast and obligatory, regardless of the stimulus type. The comparison produced a "decision code" that would identify the physically larger stimulus. While this decision code was active it caused preparation for the corresponding response to increase, but with time the decision code would diminish in strength, and hence the response preparation would diminish as well.

The process of numerical magnitude comparison was assumed to be similar to that of physical size comparison. The former produced a decision code indicating the numerically greater stimulus, and this decision code was used to select a response.

According to Foltz et al. (1984), these processes would interact to produce the Stroop effect. Following the comparison of numerical magnitudes, a response was to be selected. When the decision code for physical size was available, response preparation for the corresponding response was initiated, and the decision code would fade. Response preparation was assumed to be a cumulative, unspecified process that facilitated a particular response (Miller, 1982). On the assumption that the decision code for physical size was quickly available, congruent conditions would yield facilitation. If the decision codes for the physical and numerical comparisons were in conflict, then response competition would arise and additional processing would be required to select the appropriate decision code, thus interference would occur.

If the physical comparison stage finished before the numerical magnitude comparison begins, then there would be little response competition, and hence the Stroop effect was less likely to occur. However, if the two comparison stages occurred concurrently, then a strong response competition would arise, thus a strong Stroop effect was expected.

Foltz et al.'s (1984) account explained the Stroop effect in terms of the relative speed of the comparison processes and predicted that a stronger effect would be exerted by the comparison which was faster to accomplish on the comparison which was slower, and a weak or an absence of the effect by the latter on the former. Evidence for this prediction came from Henik and Tzelgov (1982) who, controlling for practice effect (i.e., each comparison task was not preceded by another), observed the Stroop effect (manifested as interference and facilitation) only in the numerical comparison task, not in the physical comparison task. They also reported that physical comparison was significantly (125 ms) faster than numerical comparison. The authors concluded that "Physical size as an irrelevant dimension had more influence on performance than did semantic size as an irrelevant dimension." Tzelgov et al. (1992, Experiment 3) and Girelli et al. (2000) replicated the difference in reaction times across the two comparison tasks.

Girelli et al. (2000) replicated, not only that the mean reaction time for a physical comparison task was significantly faster than that for a numerical comparison task, but also that the Stroop effect was stronger in the numerical task (both interference and facilitation were highly significant, all $p < 0.001$) than in the physical task (only interference was observed, $p < 0.050$). The findings provided support for Foltz et al.'s (1984) suggestion that "Physical size as an irrelevant dimension had more influence on performance than did semantic size as an irrelevant dimension."

However, the reversed numerical distance effect observed during physical comparisons (Girelli et al., 2000; Henik & Tzelgov, 1982) challenges Foltz et al.'s (1984) relative speed of processing account which predicted that physical size as a task-irrelevant dimension would exert a stronger influence on performance than numerical magnitude would. The reversed numerical distance effect could be viewed as an indicator of refined processing of task-irrelevant numerical magnitudes. Foltz et al.'s (1984) account would be hard-pressed to explain such a strong influence exerted by the relatively slow numerical processing on the physical comparison task.

Foltz et al.'s (1984) account would also predict a stronger Stroop effect with a stimulus type which was relatively fast to process conceptually than one which was relatively slow to process. Besner and Coltheart (1979) tested for the Stroop effect with both Arabic digits and written verbal numerals and observed the effect only with the former. Although not statistically tested, results suggest that Arabic digits (between 510 and 640 ms) were faster to process than written verbal numerals (mean reaction time = 800 ms). Thus, Besner and Coltheart's (1979) findings were consistent with Foltz et al.'s (1984) prediction. The absence of a Stroop effect with written verbal numerals could be explained in terms of a diluted effect due to the use of a fixed-pairs design. On the assumption that Arabic digits were processed significantly faster than written verbal numerals, the former would be expected to show a stronger Stroop effect, and if the use of a fixed-pairs design might have diluted the effect in both the Arabic digit and written verbal numeral conditions, it could potentially result in an absence of the Stroop effect in the latter (where the Stroop effect would be weak even if a repeated-set design were employed).

Now, consider another type of number Stroop experiment (Tzelgov et al., 1992, Experiment 1) in which subjects were required to compare the numerical magnitude and the physical size of a digit with a previously presented digit which acted as an internal standard; there was no significant difference between reaction times for physical and numerical comparisons, but the Stroop effect was observed in both tasks. Foltz et al.'s (1984) account could explain the findings in terms of an overlap between two comparison stages leading to the bi-directional Stroop effect. However, Tzelgov et al. (1992) have proposed their own explanations in terms of a distinction between intentional and autonomous information processes, based on Logan's (1988) two-process theory of skilled performance.

1.5.2 Tzelgov et al.'s (1992) Account based on Logan's (1988) Two-Process Theory of Skilled Performance

Zbrodoff and Logan (1986) referred to involuntary or unintentional processing as autonomous. Autonomy, a feature of automatic processing, has often manifested itself in Stroop tasks. According to Logan's (1988) theory, automatic performance is based on "single-step direct-access retrieval of past solutions from memory", whereas non-automatic intentional performance is based on algorithm. Performance is a result of a race between memory-based and algorithm-based processes. The likelihood that memory will win increases with practice, reflecting the automatic component of processing.

Tzelgov et al. (1992), using the number Stroop paradigm where subjects had to compare the numerical magnitude and the physical size of a digit with a previously presented digit which acted as an internal standard, observed the Stroop effect in both numerical and physical comparisons, but a numerical distance effect only in the former (Tzelgov et al., 1992, Experiments 1 & 2). The authors argued that "The absence of the distance effect in physical comparisons implies that refined numerical information does not dominate processing when subjects compare the physical size of digits to an internal standard." Hence, the findings were consistent with the proposal that numerical comparisons between two digits involve two processes, namely a large-small magnitude classification and a refined mapping that reflects the relative locations of the digits on an internal scale. (Note that the range of digits used was 1-9, where 1-4 were classified as "small" and 6-9 as "large", Tzelgov et al., 1992.) Applying Logan's (1988) theory, Tzelgov et al. (1992) argued that the first process dominated performance when numerical magnitudes were processed unintentionally (i.e., in the physical size comparison task), whereas the second process dominated performance when numerical magnitudes were processed intentionally (i.e., in the numerical magnitude comparison task). Since numerical magnitude comparisons were dominated by a slow intentional process, autonomous classification would be more likely to win the race, thus Stroop effect manifested itself in a more prominent manner; this was consistent with Henik and Tzelgov's (1982) finding that the Stroop effect (both interference and facilitation) was observed only in the numerical comparison task, not in the physical comparison task. In another experiment using a comparison type Stroop paradigm (Tzelgov et al., 1992, Experiment 3), the numerical distance effect was only observed in the numerical comparison task, but not in the physical comparison task. Tzelgov et al. (1992) concluded that "although numerical information affects physical judgments, this influence reflects mainly the large versus small classification of numerical stimuli and not their exact numerical size."

However, Logan's two-process theory could not explain the reversed numerical distance effect that numerically distant digit pairs were responded to significantly slower than close pairs (Girelli et al.; 2000; Henik & Tzelgov, 1982). Girelli et al. (2000) also observed that distant pairs exhibited a stronger Stroop effect than close pairs in numerical comparisons. The difference was more observable in physical comparisons, where only distant pairs experienced the Stroop effect and Girelli (1998) suggested that "differentiating between close pairs, being more time-demanding, would not have been completed in time for interfering with the response". Similarly, in numerical comparisons, numerically distant pairs take less time to process than numerically close pairs (the classical distance effect), so by analogy, when numerical magnitude is the task-irrelevant dimension (in physical comparisons), numerically close pairs, being more time-demanding, would not exert as much influence on physical comparisons as numerically distance pairs, hence the reversed numerical distance effect.

Girelli's (1998) view is consistent with predictions made by Algom, Dekel, and Pansky (1996) based on Melara and Mounts' (1993) notion of discriminability, which specifies the psychological difference separating two stimulus values along a dimension. In terms of the number comparison Stroop paradigm, discriminability is matched if the values along the numerical dimension (i.e., numerical magnitude) are as different psychologically as the values along the physical dimension (i.e., physical size) (Algom et al., 1996). As Melara and Mounts (1993) showed, Stroop interference is malleable, with the more discriminable dimension causing a failure of selective attention to the less discriminable dimension, but not vice versa. Thus, one would expect more interference from physically distant than close incongruent pairs on numerical comparison performance, and similarly, more disruption from numerically distant than close incongruent pairs on physical size comparisons. Indeed, Pansky and Algom (1999) found that when the task-irrelevant dimension was more discriminable, a sizeable Stroop effect affected performance on the task-relevant dimension, but when it was less discriminable, the Stroop effect was considerably weaker. These findings echo the reversed numerical distance effect observed by Henik and Tzelgov's (1982) and Girelli et al.'s (2000), on the assumption that distant pairs are more discriminable than close pairs, thus exerting a stronger size congruity effect on the task-relevant dimension.

In addition, there is another problem with Logan's (1988) model, that it fails to account for the asymmetry between facilitation and interference – the former is virtually always substantially smaller than the latter (see review, MacLeod, 1991).

1.5.3 Cohen's Parallel Distributed Processing Model

An influential model has been proposed by Cohen and colleagues (Cohen, Braver, & O'Reilly, 1996; Cohen, Dunbar, & McClelland, 1990; Cohen & Huston, 1994; Cohen & Servan-Schreiber, 1992; see also Gilbert & Shallice, 2001; Phaf, Van der Heijder, & Huston, 1990; Zhang, Zhang, & Kornblum, 1999) to explain the colour-word Stroop effect, but it can also be applied to the number Stroop effect.

By incorporating non-linear processing units in a parallel distributed processing model, Cohen et al. (1990) simulated not only the basic colour-word Stroop asymmetry that word reading interferes more with colour naming than vice versa, but also the asymmetry that interference exceeds facilitation. Cohen et al.'s (1990) model takes into account both automaticity and relative speed of processing, but the emphasis has been placed on the relative strengths (rather than speeds) of pathways in the processing system, assuming that the degree of automaticity is a function of the strength of each pathway.

Cohen et al.'s (1990) model consists of a system of connected modules. A particular process is assumed to occur via a sequence of connected modules that form a pathway. Each module consists of an ensemble of elementary processing units. Each unit is a simple information-processing device that accumulates inputs from other units and adjusts its output continuously in response to these inputs.

The model postulates that the relative strengths of two competing processes determine the patterns of the Stroop effect and the extent to which these processes are governed by attention. According to the model, interactions between processes arise when the two pathways intersect. If the patterns of activation that are generated by the processes at the point of intersection are dissimilar, then interference arises, but if they are very similar, facilitation occurs. The role of attention in the model is to select one of two competing processes according to task instructions. The stronger the pathway of a process is, the less attention it requires, and the more likely that it produces interference. The strength of a process may be increased by practice, leading to a reduction in speed of processing, and hence an increase in automaticity.

Applying Cohen et al.'s (1990) theoretical framework to the Stroop effect observed in number comparison Stroop experiments, it is reasonable to expect a stronger interference exerted on numerical comparisons by physical comparisons than the latter on the former, on the assumption that physical comparisons are more automatic (i.e., a stronger pathway strength, hence requiring less attention and processing time) than numerical comparisons.

2 Comparison Stroop Paradigm and Effects of Writing System

2.1 Introduction

The primary aim of the studies reported in Chapters 2-4 is to achieve further understanding of numerical processing through the use of a comparison Stroop paradigm. These chapters address several key issues. Firstly, findings from studies which employed the comparison Stroop paradigm are reviewed and the problems of this experimental paradigm are highlighted. Modifications of the paradigm are gradually introduced. Various groups of subject and types of stimulus have been employed in order to investigate factors which might affect performance such as writing system, relative processing speed of stimulus, and familiarity of stimulus. Both behavioural and neuroimaging techniques have been included in order to investigate the degree of numerical magnitude processing under task-relevant and -irrelevant conditions.

2.1.1 History of the Stroop Phenomenon

In 1935, Stroop published his landmark article on attention and interference, documenting the increase in reaction time in naming the colour when the coloured ink spelled the name of a conflicting colour. Named after him, the Stroop effect has never ceased to fascinate researchers since then. "Perhaps the task is seen as tapping into the primitive operations of cognition, offering clues to the fundamental process of attention" (MacLeod, 1991). The robustness of the effect has been proven to be a powerful tool to study information processing in conflicting situations over decades.

The roots of Stroop's (1935) research could be traced back half a century in the work of Cattell (1886) who reported that objects (and colours) took longer to name aloud than the corresponding words took to read aloud. For example, saying "red" to a patch of colour requires longer than saying "red" to the word *red.* Cattell (1886) explained the difference in terms of an automatic/ voluntary distinction, "This is because, in the case of words and letters, the association between the idea and name has taken place so often that the process has become automatic, whereas in the case of colours and pictures we must by a voluntary effort choose the name."

Stroop's (1935) finding could be explained in term of Cattell's (1886) automatic/ voluntary distinction. On the assumption that word reading receives more learning than colour naming, automaticity develops in the former, giving rise to the interference when required to name the colour of a conflicting colour word.

However, MacLeod (1991) argued that an all-or-none view of the automaticity explanation would be problematic because it would predict an asymmetric Stroop effect of the automatic dimension on the voluntary dimension but not vice versa. Contrary to the all-or-none view of automaticity, Glaser and Glaser (1982) observed a reversed Stroop effect in which ink colour interfered with word reading with a stimulus onset asynchrony (SOA) manipulation by increasing the frequency of congruent trials (i.e., increasing practice). Thus, others argued for a gradient of automaticity (e.g., LaBerge & Samuels, 1974; Logan, 1978; Posner & Snyder, 1975). In the case of the colour Stroop task, word reading was very automatic, and colour naming was much less automatic, hence the former would interfere with the latter, but not vice versa. However, automatic processes have to be learnt and practice would increase the degree of automaticity; increasing congruent trials in Glaser and Glaser's (1982) word reading task would result in an increase in automaticity, leading to the reversed Stroop effect. According to the graded automaticity view, if the two dimensions were of similar degree of automaticity, they would interfere with each other to more or less the same extent.

The Stroop phenomenon inspired many experiments, and numerous variants of the Stroop task have been developed. Some of these Stroop variants used numbers as stimuli, and the majority of them employed the comparison paradigm (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik & Tzelgov, 1982; Yurko & Hinrichs, 1978). The general finding – the Stroop effect – refers to the inability to ignore the numerical magnitude of the numbers when this attribute is task-irrelevant. It is important to note that this effect only emerges when children have understood the meaning (magnitude) of numbers. According to Girelli et al. (2000) who tested Italian children, this effect "was absent in first-grade children's performance, emerged in the third grade, and was highly significant in the fifth grade. In order words, the incongruity between physical and semantic information did not exert an influence on the performance before the age of 8."

2.1.2 Comparison Stroop Tasks with Numbers

In a typical comparison Stroop experiment, subjects would be asked to judge which of the number pair is larger (or smaller) either in numerical magnitude or in physical size (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik & Tzelgov, 1982; Yurko & Hinrichs, 1978). Trials may be congruent, where the numerically larger number is physically larger (e.g., 3 5); incongruent, where the numerically larger number is physically smaller (e.g., 3 5); and, in some experiments, neutral where the numbers are displayed in the same size (e.g., 3 5) for numerical comparison task and where the same numbers are displayed in different sizes (e.g., 3 3) for physical comparison task.

In Besner and Coltheart's (1979) study, subjects were asked to judge which of the number pair presented was numerically larger. The Stroop effect was observed with Arabic digits but not observed with written verbal numerals, similar to the effect (an increase in reaction times with incongruent trials compared to neutral and congruent ones) observed by Paivio (1975, Experiment 2) with pictures but not with object names in a comparison task where subjects had to judge the conceptually larger of a picture or word pair). They concluded that Arabic digits were comparable to pictorial representations, and written verbal numerals to word representations. According to Paivio (1975), "a picture has more direct access to the visual image system", whereas written verbal numerals "must be read and interpreted before they can activate the visual memory representations necessary for the size comparison", and this semantic coding process impedes the Stroop effect.

However, Foltz et al. (1984) observed the Stroop effect (manifested as both interference and facilitation) in the numerical comparison task with both digits and written verbal numerals using a repeated-set design (in which every number was paired with all other numbers). Foltz et al. (1984) attributed the absence of the Stroop effect with written verbal numerals in Besner and Coltheart's (1979) study to their use of a fixed-pairs design (in which each number was not paired with all other numbers, and number pairs were presented repeatedly), which might have allowed responses to be made upon memory of previous trials rather than upon comparison per se.

In both Besner and Coltheart's (1979) and Foltz et al.'s (1984) studies, the physical size of the numbers was manipulated as the task-irrelevant dimension and the actual comparison was based on their numerical magnitudes. Henik and Tzelgov (1982) manipulated physical size as both the task-relevant and -irrelevant dimensions. They reported that physical comparisons were significantly (125 ms) faster than numerical comparisons (replicated by Tzelgov et al., 1992) and that the Stroop effect was only observed with numerical comparisons, but not physical comparisons.

However, Girelli et al. (2000) observed the Stroop effect (manifested as interference) in both numerical and physical comparison tasks, even when the order of the tasks was controlled for. Furthermore, not only did these authors report a numerical distance effect in the numerical task, they also found a reversed numerical distance effect in the physical task, i.e., "the more distant (4 unit) pairs of digits were compared faster than the less distant (2 unit) pairs." This effect was also reported in Henik and Tzelgov's (1982) study (Experiment 2). However, Rubinsten et al. (2002) failed to replicate this finding.

The incorporation of the physical comparison task is particularly important since it allows the investigation into the processing of numerical magnitude as the task-irrelevant dimension. As mentioned before, distance effects can be used as markers of semantic processing. So, the reversed numerical distance effect observed in Henik and Tzelgov's (1982) and Girelli et al.'s (2000) studies can be interpreted as evidence for autonomous processing. However, the failure to replicate this effect means that a consensus cannot be reached. The inconsistency of this effect may be attributed to the unbalanced design commonly employed in comparison Stroop experiments. In addition, since often only two levels of numerical distance were used, the observed effect was merely a difference between two conditions (namely distance and close). It is however inaccurate to implicate a linear change from such a difference. On the other hand, by incorporating more levels of distance, linear (and any higher order) trends could be established by specifying polynomial contrasts in the analysis.

In all the numerical Stroop experiments discussed above, nine numbers were used to create two levels of numerical distance (distant and close pairs), whereas a maximum of three physical sizes were used – a large size and a small size to construct congruent and incongruent trials, and a medium size one for neutral trials, hence often only one level of physical distance was used. Moreover, there has never been an attempt to match physical distance against numerical distance. This meant that the two competing dimensions were not appropriately matched, limiting the inferences about the amount of information available in the task-relevant and -irrelevant channels.

To date, there has been no experiment which employs a parametric design, where a range of physical sizes is used as well as a range of single digits. A fully parametric design would allow one to vary systematically the influence (which may be negative, i.e., interference or positive, i.e., facilitation) exerted by the task-irrelevant dimension on the relevant dimension.

Two experiments are reported in the present chapter. Both of them examine the effects of writing system. Experiment 1 employed a traditional factorial design, where the two dimensions (namely numerical magnitude and physical size) being investigated were unbalanced, whereas in Experiment 2, the first attempt was made to vary the two dimensions parametrically. This has led to the development of a more comprehensive design in Experiments 3a and 3b (see

Chapter 3) where the influence exerted by the task-irrelevant dimension on the relevant dimension were systematically varied.

2.2 Effects of Writing System

The current section focuses on the effects of writing system on the numerical Stroop phenomenon. Firstly, the literature on such effects is reviewed with regard to both colour-word Stroop and number Stroop experiments. Then, two experiments are presented examining the numerical Stroop effect across different writing scripts.

2.2.1 Colour-Word Stroop Experiments

The colour-word Stroop effect, first documented by Stroop (1935), was replicated with colour words in other languages, e.g., Spanish (Dyer, 1971), Hungarian, French, and German (Preston & Lambert, 1969), and Chinese (Biederman & Tsao, 1979)

When Chinese-English bilingual subjects were required to name the colour of characters which represented conflicting colour words (in Chinese), they showed markedly greater interference than did English speaking subjects performing an English version of the same task (Biederman & Tsao, 1979). The authors did not attribute this effect to bilingualism among the Chinese subjects since bilingual speakers of other languages showed a weaker Stroop effect than monolingual controls (Dyer, 1971; Preston & Lambert, 1969).

Biederman and Tsao (1979) put forward an alternative, orthographic-specific explanation for the above findings, "that Chinese characters permit more direct access to meaning than an English word" which makes Chinese "much more potent than English orthography in producing Stroop-test interference". English, being an alphabetic system, is characterised by spelling-to-sound rules which allow the reader to derive the pronunciation of a word directly from the orthography. Thus, speakers are able to pronounce nonsense words insofar as their construction obeys orthographic principles (e.g., "slork", "gluck", or "vernalit"). Such, in general, is not the case with non-phonetic (e.g., logographic) representation systems. In Chinese, for example, the sound of a word and its form are for the most part arbitrarily paired. Therefore, Chinese readers must learn the pronunciation of several thousand logographs without the aid of an orthographic rule system. Based on these differences, Biederman and Tsao (1979) suggested that the application of a system where the names were directly associated to the configural appearance of stimuli, as in Chinese, would result in a stronger Stroop effect than the application of an abstract rule system, as in English.

Contrary to Biederman and Tsao's (1979) assumption that reading Chinese is strictly a print-to-meaning process, Leong (1997) argued that Chinese characters map onto both meaning and phonology. Hung and Tzeng (1981) reported that about four in five Chinese characters have a radical which indicates the pronunciation. However, they are unreliable, so only two in five Chinese logographs, if unfamiliar, can be pronounced from the phonetic radicals. Others reported different estimates on the validity of the phonetic components (or phonetics), ranging from 27% (Fan, Gao, & Ao, 1984) to 48% (Zhou, 1978), depending on specific calculation assumptions.

More recent research is also in disagreement with Biederman and Tsao's (1979) view. Phonological processing of Chinese characters has been observed very early (Cheng & Shih, 1988; Perfetti & Zhang, 1991; Tan & Peng, 1991), perhaps earlier than meaning processing in various tasks (Perfetti & Zhang, 1995; Tan, Hoosain, & Peng, 1995; Tan, Hoosain, & Siok, 1996). For example, in a backward masking paradigm, Tan et al. (1996) found that a character's phonology is accessed earlier than its semantics. Using meaning and homophone judgement tasks, Perfetti and Zhang (1995) observed that phonological interference occurred before semantic interference. These findings challenge the strictly print-to-meaning view of reading Chinese, and constitute, instead, evidence for a phonology-before-meaning hypothesis. Perfetti, Zhang, and Berent (1992) argued for a universal phonological principle which suggests that phonological processing is a characteristic of human language processing rather than a by-product of the writing system and that phonological processes occur as part of reading in all writing systems.

Furthermore, research on brain injuries provides support to the idea that Chinese logographs can be read phonologically to a certain extent. For example, in Yin and Butterworth's (1992) study, Chinese surface dyslexics demonstrated the ability to use the phonetic radicals, where available, in reading characters aloud, making no semantic errors. The opposite effect – a loss in ability to use phonetic radicals to identify characters – was observed in Chinese alexics (Yin & Butterworth, 1998).

Yin and Butterworth (1992) argued that a two-routine model applies to the reading of both alphabetic and non-alphabetic scripts. Models, such as Coltheart's (1978) dual-route reading system, claim that skilled adults can read words aloud in two fundamentally distinct ways: (1) by means of a lexical procedure that associates a letter string (or a character in the case of Chinese) as a whole with a meaning and a whole-word (or whole-character) pronunciation, and (2) by means of a sub-lexical procedure that associates single letters or small groups of letters (or radicals in the case of Chinese) with separate syllables or phonemes (or phonetic radicals or components in the case of Chinese), which then have to be assembled to yield the pronunciation of the whole word (or character).

The pattern of Chinese surface dyslexics is consistent with that of surface dyslexics of non-alphabetic scripts such as English – a deficit to the whole word routine where individuals have particular difficulty with irregularly spelled words like "pint" (e.g., Coltheart, 1982; see also Shallice, Warrington, & McCarthy, 1983), but the routine for mapping letters onto phoneme (or phonetic radicals or components in the case of Chinese) is intact. Similarly, the pattern of Chinese alexics is consistent with that of individuals with phonological alexia (Beauvois & Derouesne, 1979) of non-alphabetic scripts such as English – the sub-lexical route is impaired where individuals are unable to read novel letter strings (Marshall & Newcombe, 1973), while mapping from whole words is relatively intact. Yin and Butterworth (1992) concluded that "Despite the very different ways in which Chinese and alphabetic writing systems represent their languages, the reading processes that decode them appear to have the same broad underlying cognitive architecture."

In the dual-routine model described above, the lexical route implies a direct link between orthography and semantics. This view is supported by Tan and Perfetti's (1997) finding that synonyms are better than homophones of synonyms in priming the name of a target character. These authors also argued that the activation of phonological forms causes meaning to be activated, but not necessarily accessed.

The general consensus of recent studies points to the direction that reading processes are in general the same across writing systems. So, what gave rise to difference in the manifestation of the Stroop effect across stimuli of different languages? One possible explanation for the stronger Stroop effect observed with Chinese characters compared to stimuli of alphabetic scripts is in terms of Cohen et al.'s (1990) idea of the strength of a process. According to these authors, the stronger the pathway of a process is, the less attention it requires, and the more likely that it produces interference. Perhaps the direct orthography-to-semantics route is stronger and more preferable during a Stroop task than the indirect route via phonology in Chinese than in alphabetic scripts, therefore inducing more interference on the relevant task, colour naming.

2.2.2 Number Stroop Experiments

In comparison Stroop experiments with numbers, subjects are typically asked to judge which of the pair is larger (or smaller) numerically and physically. Trials may be congruent, neutral, or incongruent. The general findings of comparison Stroop tasks are as follows. The Stroop effect manifests as interference (indicated by an increase in mean reaction time/ error rate across neutral and incongruent trials) and/ or facilitation (indicated by a reduction in mean reaction time/ error rate across neutral and congruent trials) (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik & Tzelgov, 1982; Tzelgov et al., 1992). Notably, facilitation is virtually always substantially smaller than interference (see review, MacLeod, 1991).

Besner and Coltheart (1979) observed the Stroop effect with Arabic digits but not with written verbal numerals, consistent with the Stroop effect observed by Paivio (1975) with memorial pictures, but not object names, in a task where subjects had to judge the larger of the pair conceptually. Besner and Coltheart (1979) concluded that Arabic digits were comparable to pictorial representations, whereas written verbal numerals were comparable to word representations. According to Paivio (1975), "a picture has more direct access to the visual image system", whereas written verbal numerals "must be read and interpreted before they can activate the visual memory representations necessary for the size comparison", and this semantic coding process impedes the Stroop effect.

However, Foltz et al. (1984) observed the Stroop effect (manifested as both interference and facilitation) with both Arabic digits and written verbal numerals using a repeated-set design (in which every number was paired with all other numbers). Foltz et al. (1984) attributed the absence of the Stroop effect with written vernal numbers in Besner and Coltheart's (1979) study to their use of a fixed-pairs design (in which each number was not paired with all other numbers, and number pairs were presented repeatedly), which might have allowed responses to be made upon memory of previous trials rather than upon comparison per se.

Foltz et al.'s (1984) relative speed of processing account predicted a stronger Stroop effect with a stimulus type which was relatively fast to process conceptually than one which was relatively slow to process. Although not statistically tested, results suggested that Arabic digits (between 510 and 640 ms, Besner & Coltheart, 1979; mean 585 ms, Foltz et al., 1990) were faster to process than written verbal numerals (mean 800 ms, Besner & Coltheart, 1979; mean 762 ms, Foltz et al., 1990). The findings that, when a fixed-pair design was employed, the Stroop effect was only observed with Arabic digits and not with written verbal numerals (Besner & Coltheart, 1979), whereas when a repeated-set design was used, the Stroop effect was observed in both types of stimulus (Foltz et al., 1990) were consistent with Foltz et al.'s (1990) account – the absence of a Stroop effect with written verbal numerals can be explained in terms of a diluted effect due to the use of a fixed-pairs design. On the assumption that Arabic digits are processed significantly faster than written verbal numerals, the former will be more susceptible to a Stroop effect, hence the use of a fixed-pairs design will be less capable of diluting the effect.

Very few attempts have been made to investigate the effects of writing system on the Stroop phenomenon (Takahashi & Green, 1983; Tzeng & Wang, 1983; Vaid, 1985). Unfortunately, these early studies were poorly designed, leading to unwarranted findings.

Tzeng and Wang (1983) tested two groups of bilingual subjects – Chinese native speakers and Spanish native speakers, who had learned English as young adults – with a numerical comparison task. The Chinese-English bilingual readers showed significant interference with Arabic digits, and both Chinese and English written verbal numerals. However, the Spanish-English bilingual readers showed interference only with Arabic digits, but not Spanish or English written verbal numerals. The authors concluded that "Therefore the interference observed with English words for the Chinese-English bilingual readers was not due to their latter learning of that language, since we did not observe a similar effect with Spanish-English readers. The remaining hypothesis, then, was that the Chinese-English subjects had transferred their reading habits from logographs to English words." However, the authors provided no statistical support for their claims, largely discrediting their conclusion.

Takahashi and Green (1983) also employed a numerical comparison task to test for differences in Japanese subjects with the two Japanese scripts: Kanji and Kana. On the basis of Besner and Coltheart's (1979) finding that the Stroop effect was only observed with Arabic digits (an ideographic script) but not written verbal numerals (an alphabetic script which is tied to the spoken form), Takahashi and Green (1983) drew a parallel prediction that the Stroop effect would only be observed with Kanji written verbal numerals (an ideographic script) but not with Kana numeral equivalents (a syllabic, non-ideographic script).

Indeed, when controlling for order effect (i.e., a condition was not preceded by another), the Stroop effect (manifested as interference and facilitation) was only observed with Kanji written verbal numerals but not with the Kana numeral equivalents. However, when the Kana condition was preceded by the Kanji one, interference was observed with the Kana numeral equivalents. Similar to the conclusion provided Tzeng and Wang (1983), Takahashi and Green (1983) suggested that the subjects could be "treating the script ideographically" as if they had transferred their reading habits from Kanji reading.

In addition, Takahashi and Green (1983) also observed the numerical distance effect with both Kanji and Kana stimuli. They noted the difference between the reaction time functions for numerical distance between the two scripts, "slightly concave for Kanji and convex for Kana". Based on this difference, the authors argued "that Kana characters either access a different numerical representation from Kanji characters, or invoke different procedures for comparing the two numbers." However, the authors did not further specify the differences.

There is an alternative account for the different performance with the two scripts. Japanese numerals, unlike other words, can only be written in Kanji. Thus, numerals written in Kana are, by definition, non-words. This results in a familiarity discrepancy between the two types of stimulus. The absence of a Stroop effect with Kana written numeral equivalents might have resulted from a lack of experience; such unfamiliar stimuli would be harder for the subjects to process, hence more attention would be required for the comparison task. The increase in attention might have led to a better ability to filter out the task-irrelevant information (physical size), thus impeding any potential Stroop effect.

Vaid (1985) also examined the effect of writing system on the Stroop phenomenon. Again, only a numerical comparison paradigm was employed. Vaid (1985) distinguished between three types of writing systems: (1) logographic (or ideographic) scripts, such as Chinese, map onto speech at the level of the morpheme, (2) alphabetic scripts map onto speech either at the phonemic level, as in Serbo-Croatian, or at the morphophonemic level, as in English, and (3) syllabic scripts, such as the Japanese Kana and the Indian

Devanagari, map onto speech at level of the syllable. Vaid (1985) compared three scripts: Arabic digits (logographic), English written verbal numerals (alphabetic), and Hindi written verbal numerals (syllabic). Based on the collected findings from Besner and Coltheart's (1979), Takahashi and Green (1983), Tzeng and Wang (1983), Vaid (1985) formulated the hypothesis that performance in the Stroop task would depend on the phonological transparency of the writing script, in the direction that the less transparent the script, the greater the Stroop effect. In particular, the author predicted that ideographic scripts, such as Arabic digits, "encourage the use of a direct visual coding" which "is detrimental to performance", hence these stimuli would exhibit a very strong Stroop effect. The author also predicted that English written verbal numerals (alphabetic script) would exhibit a weaker Stroop effect on the basis that this type of writing script encourages phonological recoding, and that Hindi written verbal numerals (syllabic script) would experience the weakest or an absence of a Stroop effect due to its high phonological transparency (i.e., high reliance on phonological recoding).

Consistent with Vaid's (1985) predictions, the Stroop effect was observed with Arabic digits and English written verbal numerals, but not with Hindi written verbal numerals. This reinforces the hypothesis that "phonological encoding was a consistent strategy used by subjects when making numerical size comparisons in Hindi." The author also commented that "physical size cues influenced judgments of numerical size to about the same extent in the numeral and English word conditions. This would suggest that a visual representation of the input was used in performance on the task." However, Vaid's (1985) claim that the Stroop effect was "about the same extent" with Arabic digits and English written verbal numerals has been undermined by the lack of statistical evidence. Furthermore, the experimental design only employed congruent and incongruent but not neutral trials, consequently interference and facilitation could not be disentangled. The poor experimental design accompanied by insufficient statistical analyses has hugely discredited Vaid's (1985) findings.

In all three experiments described above (Takahashi & Green, 1983; Tzeng & Wang, 1983; Vaid, 1985), only the numerical comparison task was employed, but not the physical comparison task. As mentioned before, the latter has the advantage of allowing investigation into the autonomous property of numerical magnitude processing. Hence, a comprehensive design should always include both tasks.

In the following sections, two experiments with different groups of bilingual subjects are presented, focusing on an investigation into the effects of writing system. They aimed to battle some of the flaws in previous studies. Attempts were made to control for subjects' fluency in the languages that they spoke. More importantly, physical comparisons were incorporated in order to gain understanding into the processing of numerical magnitude as the task-irrelevant dimension.

Experiment 1 investigated the processing of written verbal numerals in Chinese and English. Chinese-English bilinguals possessing similar fluency levels in the two languages were recruited.

In Experiment 2, two Japanese writing scripts, Kanji and Kana, were investigated. "Kanji" in Chinese is 漢字, meaning "Chinese character" in both languages. In fact, Kanji and Chinese share the same written verbal numerals. "Kana" is a collective term for two Japanese scripts, Hiragana which is used for native words and Katakana which is used for foreign words. The major difference between Kanji and Kana is that the former is ideographic and the latter is syllabic. Interestingly, Japanese verbal numerals are usually written in Kanji and rarely in Kana. This experiment was designed to investigate the processing of Kanji written verbal numerals and their unfamiliar Kana equivalents. A first attempt was made to develop a parametric design for the comparison Stroop paradigm.

2.3 *Experiment 1: Comparing Numerical Magnitudes in Chinese and in English*

2.3.1 Methods

2.3.1.1 Tasks

The two tasks were numerical magnitude comparison and physical size comparison. Subjects had to select, via a key response (by pressing the left "F" key with their left index finger or the right "J" key with their right index on a qwerty keyboard) the larger number of the pair in numerical magnitude or in physical size accordingly. Their reaction times and responses were recorded.

The tasks were computer-based. A program written in MATLAB/ Cogent was used. Reaction time (time, in milliseconds, for the subject to make a response after the presentation of a pair of stimuli) and the response (left or right key press) were recorded for each of the 120 trials in each task.

2.3.1.2 Stimuli

There were three types of stimulus: Arabic digits, Chinese written verbal numerals, and English written verbal numerals (see Table 2.1).

Table 2.1 Stimuli – Arabic digits, Chinese and English written verbal numerals – in font sizes 8 and 12 respectively (Experiment 1)

Arabic Digits		Chinese Written Verbal Numerals		English Written Verbal Numerals	
1	1	一	一	ONE	ONE
2	2	二	二	TWO	TWO
3	3	三	三	THREE	THREE
4	4	四	四	FOUR	FOUR
6	6	六	六	SIX	SIX
7	7	七	七	SEVEN	SEVEN
8	8	八	八	EIGHT	EIGHT
9	9	九	九	NINE	NINE

The font styles used in the present experiment were Times New Roman for Arabic and English stimuli, and 全真楷書 for Chinese stimuli.

A pilot study was carried out prior to the current experiment and it was observed that when physical sizes were similar, physical comparisons were particularly difficult with Chinese written verbal numerals (reflected by an unexpectedly high mean reaction time of > 600 ms). This was due to the fact that Chinese written verbal numerals have a large variation in terms of maximum height for any given font size (for instance, the numbers in the pair (1 6) in Chinese (一 六) of the same font size have different heights), unlike the other stimulus types, namely English written verbal numerals (e.g., ONE SIX) and Arabic digits (e.g., 1 6) which have constant maximum height for any given font size. The difficult physical comparisons with Chinese written verbal numerals led to an exaggerated Stroop effect in the pilot study, so efforts were made in the current experiment to ensure that the physical discriminability was similar across the three stimulus types. It must be noted that English written verbal numerals have a large variation in terms of maximum width compared to other stimulus types. Table 2.2 shows the physical sizes of the stimuli used in the current experiment.

Table 2.2 Maximum heights (cm) of different stimulus types (Experiment 1)

Stimulus Type	Size		
	Small	Large	Neutral
Arabic	0.6	1.4	1.0
English	0.6	1.3	1.0
Chinese	0.8	1.5	1.1

In each trial, a pair of stimuli, 14.50 cm apart, appeared in the middle of the computer screen for 1000 ms, and there was an interval of 1000 ms before the subsequent trial. Stimuli appeared in white on a black background in the middle of the 15-inch computer screen.

The experimental design followed that of Girelli et al.'s (2000). The different types of pair of numbers were: numerically close pairs, numerically distant pairs, and neutral pairs. In close pairs, the numbers were either small (1, 2, 3, or 4) or large (6, 7, 8, or 9), and they were separated by a numerical distance of 1. The 12 pairs were 1 2, 2 1, 2 3, 3 2, 3 4, 4 3, 6 7, 7 6, 7 8, 8 7, 8 9, 9 8. "5" was not included because it has equal associative strength with "small" and "large", i.e., it would be classified as "smaller" half of the time and "larger" half of the time when compared with every other digit an equal number of times (Tzelgov et al., 1992). In numerically distant pairs, one number was small (1, 2, 3, or 4) and the other was large (6, 7, 8, or 9), and they were separated by a numerical distance of 5. The eight pairs were 1 6, 6 1, 2 7, 7 2, 3 8, 8 3, 4 9, 9 4. In the numerical comparison task, neutral pairs were identical to the congruent and incongruent pairs (e.g., 20 for each level of congruity), where the two numbers of each pair were displayed in the same physical size. Neutral pairs in the physical comparison task were 1 1, 2 2, 3 3, 4 4, 6 6, 7 7, 8 8, 9 9, where the two numbers of each pair were displayed in different physical sizes.

There were three levels of congruity: (a) congruent, when the numerically larger number was physically larger (e.g., 2 7); (b) incongruent, when the numerically larger number was physically smaller (e.g., 2 7); (c) neutral, when the numbers were displayed in the same size (in numerical comparison; e.g., 2 7) or the same numbers were displayed in different sizes (in physical comparison; e.g., 2 2). In the numerical comparison task, each single pair was presented twice in each condition, making a total of 120 trials. In the physical comparison task, each pair was presented twice in congruent and incongruent conditions, while neutral pairs were presented five times each (i.e., 40 neutral trials, the equivalent of one third of the experimental trials), for a total of 120 trials. In all conditions in which physical sizes varied, each trial was presented once with the physically larger digit on the left and once on the right. In the neutral condition of the physical comparison task, the frequency of the relative position of the physically larger digit was counterbalanced across the trials. Thus, in both tasks, the position of the correct answer was balanced across trials.

Stimuli were presented in pseudorandom order with the following constraints to avoid carryover effects (Girelli et al., 2000): (1) the same number did not appear in consecutive trials, (2) the correct answer did not appear on the same side (left or right) for more than three consecutive trials, (3) the numerical distance of the stimuli (i.e., close and distant) was not the same for more than three consecutive trials, and (4) the experimental condition (i.e., congruent, incongruent, and neutral) was not the same for more than two consecutive trials.

2.3.1.3 Procedures

A brief introduction about the whole experiment was given at the beginning. Instructions for different tasks were given at the start of each.

Practice trials were given to the subjects to ensure that they understood the tasks. The pairs of numbers used in the practice trials did not occur in any of the experimental trials; this was done to eliminate any possible priming effect.

Half of the subjects in each of the two groups (the bilingual and the control groups) started with a numerical task and the other half started with a physical task, and consecutive blocks were never the same type of comparison. Subjects always performed the comparison tasks of the same stimulus type consecutively (e.g., numerical task followed by physical task with English written verbal numerals, numerical task followed by physical task with Arabic digits). Changing stimulus type might cause confusion to subjects, so it was decided that one task would be followed by another of the same stimulus type. The order in which the three stimulus types were given to the subjects was counterbalanced.

2.3.1.4 Subjects

The subjects were students and graduates at the University of London. There were 24 right-handed subjects (8 males and 16 females), aged from 19 to 35 (mean 23.9 years, standard deviation = 3.8 years). Half of the subjects were Chinese-English bilingual speakers and the other half were native English speakers (control subjects) who had received all their education in England.

The bilingual subjects' first language was Chinese (Cantonese) and they started learning English as a second language between the ages 3 and 9 (mean 5.3 years). All of them were born in Hong Kong and had between 9 and 19 years (mean 12.2 years) of formal (pre-primary/ kindergarten, primary, and some or all secondary/ tertiary) education in Hong Kong, where the principal language used was Chinese (Cantonese). They all started learning mathematics in Chinese at kindergarten. During secondary education, mathematics was taught in English.

All subjects had normal or corrected-to-normal eyesight.

2.3.2 Results

ANOVAs were used to analyse the mean percentage errors and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

2.3.2.1 Mean Error Rates

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 1000 ms after the stimulus offset.

Three subjects who made more than 10% errors in either one task were removed. A 2 x 2 x 3 x 2 mixed design ANOVA was conducted on mean error rates comparing the bilingual and the control subjects. The four factors were: subject (bilingual and control subjects), task (numerical and physical comparisons), stimulus (Arabic digits and English written verbal numerals), and congruity (congruent, neutral and incongruent). Subject was the only between-subjects factor. The ANOVA revealed a significant main effect of task ($F_{(1,19)} = 47.98$, $p < 0.001$), a significant main effect of stimulus ($F_{(1,19)} = 17.48$, $p < 0.050$), and a significant main effect of congruity ($F_{(2,38)} = 19.48$, $p < 0.001$). The main effect of subject was not significant ($F_{(1,19)} < 1$, *n.s.*), and the factor subject did not interact significantly with task or with congruity (both *n.s.*). However, the stimulus x subject interaction was significant ($F_{(1,19)} = 5.52$, $p < 0.050$). The factor congruity also interacted significantly with other factors: task x congruity ($F_{(2,38)} = 8.51$, $p \leq 0.001$), and stimulus x congruity ($F_{(1,23)} = 5.06$, $p < 0.050$). All other interactions were not significant (all *n.s.*).

Due to the various significant interactions, further analyses were conducted separately for the two groups of subjects.

2.3.2.1.1 Control Subjects

A 2 x 2 x 3 ANOVA was conducted on mean error rates. The factors were task (numerical and physical), stimulus (Arabic digits and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of task ($F_{(1,9)} = 15.47$, $p < 0.005$). The main effect of stimulus was non-significant ($F_{(1,9)} = 1.95$, *n.s.*). On the other hand, there was a significant main effect of congruity ($F_{(2,18)} = 8.65$, $p < 0.005$), and the factor congruity interacted significantly with other factors: task x congruity interaction ($F_{(2,18)} = 3.97, p < 0.050$), and stimulus x congruity interaction ($F_{(2,18)} = 6.32, p < 0.050$). All other interactions were non-significant (all *n.s.*).

One-way ANOVAs were performed at each level of task for the two stimulus types separately.

For Arabic digits, a significant main effect of congruity was observed during numerical comparisons ($F_{(2,18)} = 7.33, p \leq 0.005$), but within-subjects contrasts revealed no significant difference between congruent and neutral trials, and between neutral and incongruent trials (both *n.s.*). The mean error rates were 0.25%, 2.75%, and 6.75% respectively (see Figure 2.1). The main effect of congruity was non-significant during physical comparisons ($F_{(2,18)} = 3.20$, *n.s.*). The mean error rates were 0.00%, 0.50%, and 2.50% for congruent, neutral, and incongruent trials respectively (see Figure 2.1).

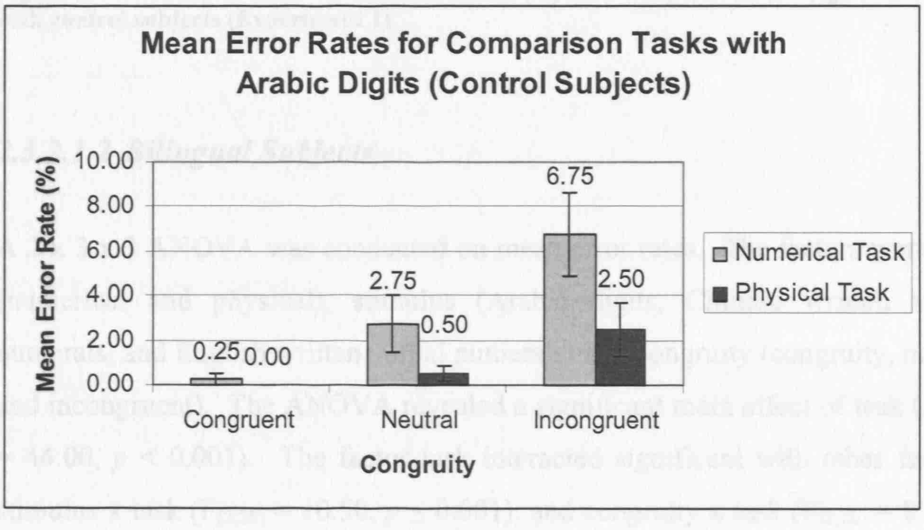


Figure 2.1 Mean error rates for numerical and physical comparisons of Arabic digits with control subjects (Experiment 1)

For English written verbal numerals, a significant main effect of congruity was observed during numerical comparisons ($F_{(2,18)} = 5.69, p < 0.050$). Within-subjects contrasts revealed a significant difference between congruent and neutral trials ($F_{(1,9)} = 8.31, p < 0.050$), but not between neutral and incongruent trials ($F_{(1,9)} < 1$, *n.s.*). The mean error rates were 2.75%, 5.75%, and 5.25%

respectively (see Figure 2.2). The main effect of congruity was non-significant during physical comparisons ($F_{(2,18)} = 1.33$, *n.s.*). The mean error rates were 0.50%, 1.25%, and 0.50% for congruent, neutral, and incongruent trials respectively (see Figure 2.2).

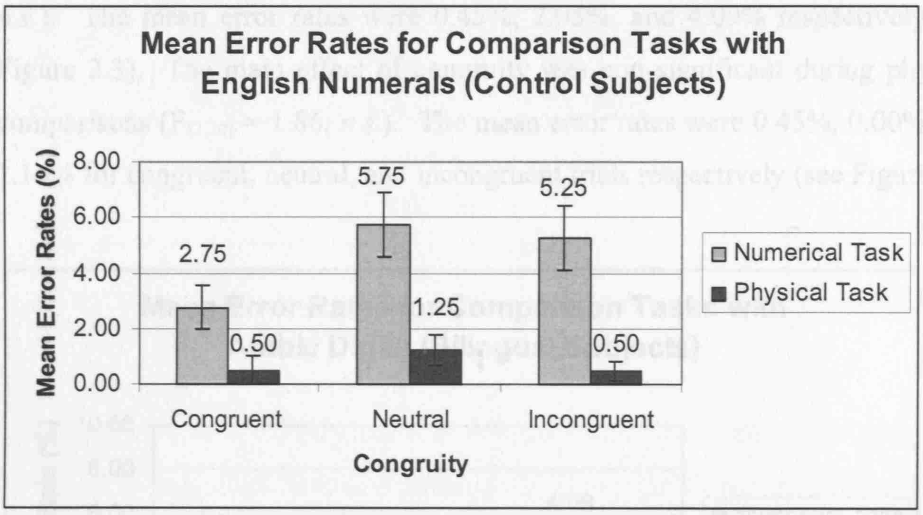


Figure 2.2 Mean error rates for numerical and physical comparisons of English numerals with control subjects (Experiment 1)

2.3.2.1.2 Bilingual Subjects

A 2 x 3 x 3 ANOVA was conducted on mean error rates. The factors were task (numerical and physical), stimulus (Arabic digits, Chinese written verbal numerals, and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of task ($F_{(1,10)} = 44.00$, $p < 0.001$). The factor task interacted significant with other factors: stimulus x task ($F_{(2,20)} = 10.50$, $p \leq 0.001$), and congruity x task ($F_{(2,20)} = 8.86$, $p < 0.005$). There was also a significant main effect of stimulus ($F_{(2,20)} = 14.72$, $p < 0.001$), and a significant main effect of congruity ($F_{(1,12)} = 17.49$, $p \leq 0.001$), but the stimulus x congruity interaction was non-significant ($F_{(2,17)} = 3.26$, *n.s.*). The three-way interaction was also non-significant ($F_{(2,19)} = 1.42$, *n.s.*).

One-way ANOVAs were performed at each level of task for the two stimulus types separately.

For Arabic digits, a significant main effect of congruity was observed during numerical comparisons ($F_{(2,20)} = 6.17$, $p < 0.050$). Within-subjects contrasts revealed a significant difference between congruent and neutral trials ($F_{(1,10)} = 9.80$, $p < 0.050$), but not between neutral and incongruent trials ($F_{(1,10)} = 2.49$, *n.s.*). The mean error rates were 0.45%, 2.05%, and 4.09% respectively (see Figure 2.3). The main effect of congruity was non-significant during physical comparisons ($F_{(2,20)} = 1.86$, *n.s.*). The mean error rates were 0.45%, 0.00%, and 1.14% for congruent, neutral, and incongruent trials respectively (see Figure 2.3).

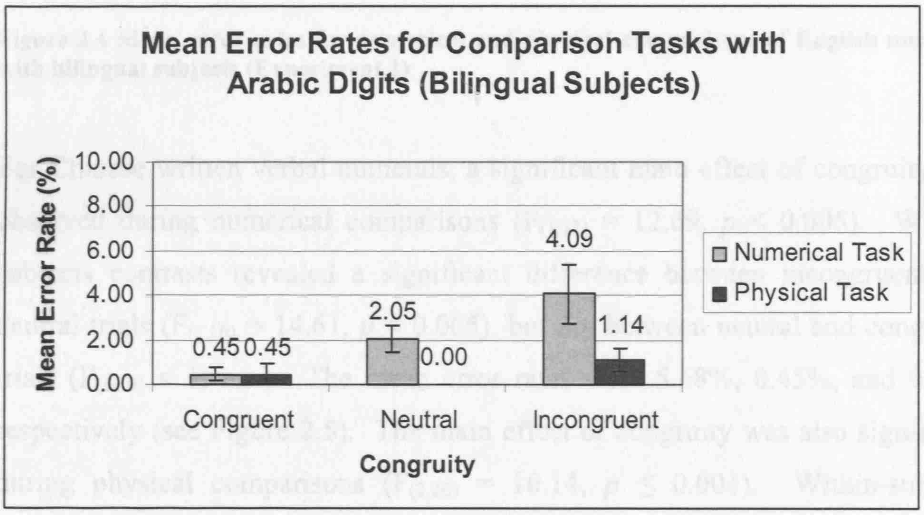


Figure 2.3 Mean error rates for numerical and physical comparisons of Arabic digits with bilingual subjects (Experiment 1)

For English written verbal numerals, there was a non-significant main effect of congruity during numerical comparisons ($F_{(2,20)} = 2.18$, *n.s.*). The mean error rates were 4.32%, 7.05%, and 6.59% for congruent, neutral, and incongruent trials respectively (see Figure 2.4). The main effect of congruity was also non-significant during physical comparisons ($F_{(1,11)} = 4.30$, *n.s.*). The mean error rates were 0.00%, 0.23%, and 1.59% for congruent, neutral, and incongruent trials respectively (see Figure 2.4).

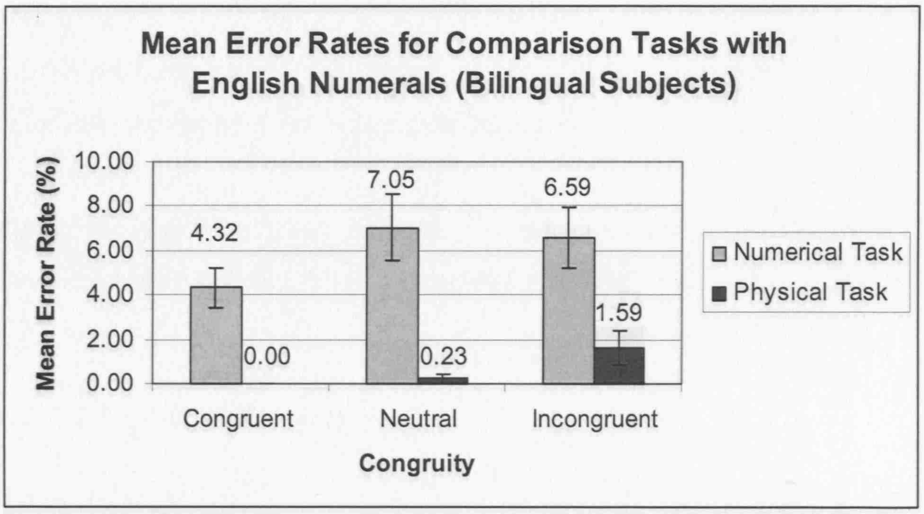


Figure 2.4 Mean error rates for numerical and physical comparisons of English numerals with bilingual subjects (Experiment 1)

For Chinese written verbal numerals, a significant main effect of congruity was observed during numerical comparisons ($F_{(1,12)} = 12.69$, $p < 0.005$). Within-subjects contrasts revealed a significant difference between incongruent and neutral trials ($F_{(1,10)} = 14.61$, $p < 0.005$), but not between neutral and congruent trials ($F_{(1,10)} < 1$, *n.s.*). The mean error rates were 5.68%, 0.45%, and 0.45% respectively (see Figure 2.5). The main effect of congruity was also significant during physical comparisons ($F_{(2,20)} = 10.14$, $p \leq 0.001$). Within-subjects contrasts revealed a significant difference between incongruent and neutral trials ($F_{(1,10)} = 10.14$, $p < 0.050$). The mean error rates were 2.73%, 0.00%, and 0.00% for incongruent, neutral, and congruent trials respectively (see Figure 2.5).

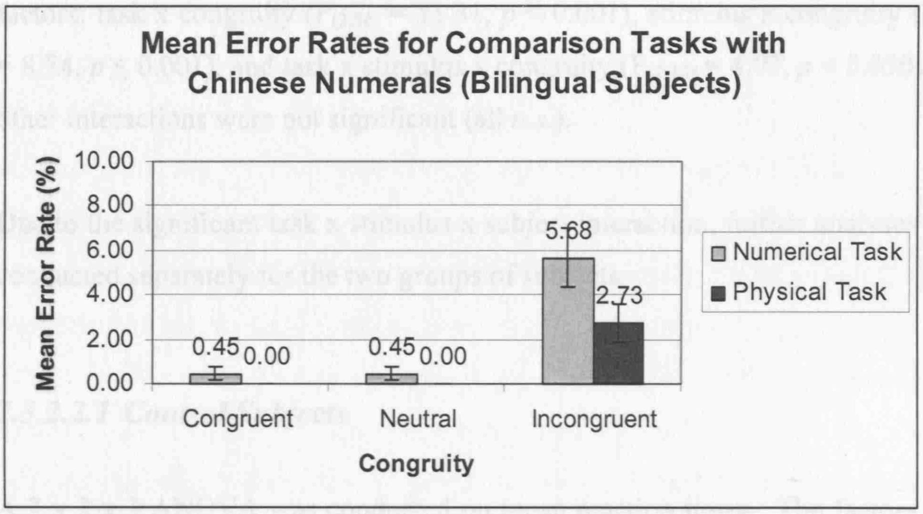


Figure 2.5 Mean error rates for numerical and physical comparisons of Chinese numerals with bilingual subjects (Experiment 1)

2.3.2.2 Mean Reaction Times

Errors and reaction time outliers were removed before further analyses. Outliers included missed trials and values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile.

A 2 x 2 x 3 x 2 mixed design ANOVA was conducted on mean reaction times comparing the bilingual and the control subjects. The four factors were: subject (bilingual and control subjects), task (numerical and physical comparisons), stimulus (Arabic digits and English written verbal numerals), and congruity (congruent, neutral and incongruent). Subject was the only between-subjects factor. The ANOVA revealed a significant main effect of task ($F_{(1,19)} = 344.48$, $p < 0.001$), a significant main effect of stimulus ($F_{(1,19)} = 82.58$, $p < 0.001$), and a significant main effect of congruity ($F_{(2,38)} = 45.96$, $p < 0.001$). The main effect of subject was not significant ($F_{(1,19)} < 1$, *n.s.*), and the factor subject did not interact significantly with task, stimulus, or congruity (all *n.s.*). However, the task x stimulus x subject interaction was significant ($F_{(1,19)} = 5.12$, $p < 0.050$), which resulted primarily from the significant task x stimulus interaction ($F_{(1,19)} = 250.25$, $p < 0.001$). The factor congruity also interacted significantly with other factors: task x congruity ($F_{(2,38)} = 11.31, p < 0.001$), stimulus x congruity ($F_{(2,38)} = 8.74, p \leq 0.001$), and task x stimulus x congruity ($F_{(2,38)} = 4.07, p < 0.050$). All other interactions were not significant (all *n.s.*).

Due to the significant task x stimulus x subject interaction, further analyses were conducted separately for the two groups of subjects.

2.3.2.2.1 Control Subjects

A 2 x 2 x 3 ANOVA was conducted on mean reaction times. The factors were task (numerical and physical), stimulus (Arabic digits and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of task ($F_{(1,9)} = 261.25$, $p < 0.001$; mean reaction times were 638 ms and 412 ms for numerical and physical comparisons respectively); a significant main effect of stimulus ($F_{(1,9)} = 39.20$, $p < 0.001$; mean reaction times were 491 ms and 559 ms for Arabic digits and English written verbal numerals respectively), a significant main effect of congruity ($F_{(2,18)} = 15.73$, $p < 0.001$), a significant task x stimulus interaction ($F_{(1,9)} = 163.67, p < 0.001$), a significant task x congruity interaction ($F_{(2,18)} = 3.96, p < 0.050$), a significant stimulus x congruity interaction ($F_{(2,18)} = 3.74, p < 0.050$), and a non-significant task x stimulus x congruity interaction ($F_{(2,18)} = 2.09$, *n.s.*).

One-way ANOVAs were performed on mean reaction times at each level of stimulus. For Arabic digits, physical comparisons were significantly faster than numerical comparisons ($F_{(1,9)} = 81.65, p < 0.001$; mean reaction times were 423 ms and 558 ms respectively). For English written verbal numerals, physical comparisons were significantly faster than numerical comparisons ($F_{(1,9)} = 370.83, p < 0.001$; mean reaction times were 400 ms and 718 ms respectively).

Considering only the numerical task, a 2 x 3 ANOVA was conducted on mean reaction times. The factors were stimulus (Arabic digits and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of stimulus ($F_{(1,9)} = 90.85, p < 0.001$), a significant main effect of congruity ($F_{(2,18)} = 10.27$, $p \leq 0.001$), and a significant stimulus x congruity interaction ($F_{(2,18)} = 3.66$, $p < 0.050$).

Further analyses revealed that Arabic digits were responded to significantly faster than English written verbal numerals ($t_{(9)} = 9.53$, $p < 0.001$); mean reaction times were 558 ms and 718 ms respectively. For Arabic digits, congruent trials were responded to significantly faster than neutral trials ($F_{(1,9)} = 10.78$, $p < 0.050$), which were in turn responded to significantly faster than incongruent trials ($F_{(1,9)} = 13.01$, $p < 0.050$). The mean reaction times were 531 ms, 561 ms, and 583 ms respectively (see Figure 2.6). On the other hand, English written verbal numerals did not show any significant difference in mean reaction time across the congruity levels (all *n.s.*). The mean reaction times were 709 ms, 717 ms, and 727 ms for congruent, neutral, and incongruent trials respectively (see Figure 2.7).

Considering only the physical task, a 2 x 3 ANOVA was conducted on mean reaction times. The factors were stimulus (Arabic digits and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of stimulus ($F_{(1,9)} = 9.60$, $p < 0.050$), a significant main effect of congruity ($F_{(2,18)} = 8.63$, $p < 0.050$), and a non-significant stimulus x congruity interaction ($F_{(2,18)} = 3.21$, *n.s.*).

Further analyses revealed that Arabic digits were responded to significantly slower than English written verbal numerals ($t_{(9)} = 3.10$, $p < 0.050$); mean reaction times were 432 ms and 400 ms respectively. Mean reaction time for congruent trials did not differ significantly from that for neutral trials ($F_{(1,9)} = 1.01$, *n.s.*), but the latter were responded to significantly faster than incongruent trials ($F_{(1,9)} = 10.39$, $p < 0.050$). The mean reaction times were 413 ms, 420 ms, and 437 ms respectively (see Figure 2.6). The non-significant stimulus x congruity interaction indicated that the same pattern applied to both stimulus types. With English written verbal numerals, the reaction times were 399 ms, 400 ms, and 400 ms for congruent, neutral, and incongruent trials respectively (see Figure 2.7).

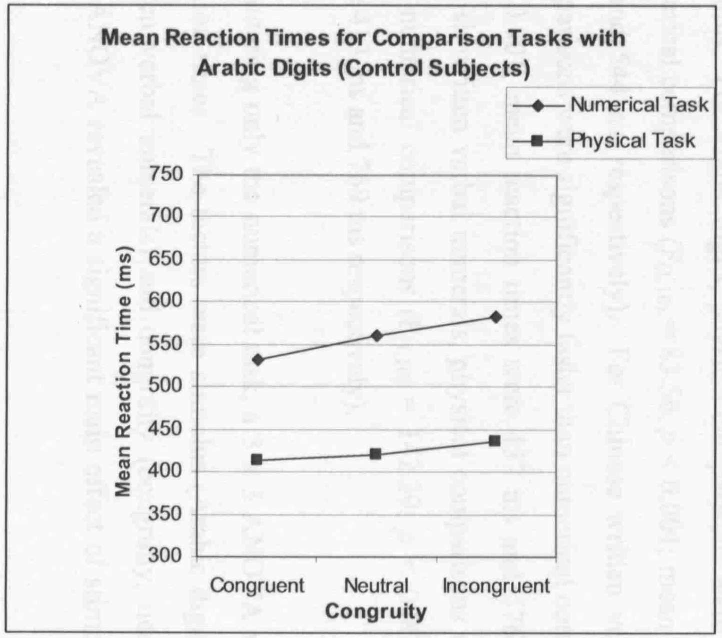


Figure 2.6 Mean reaction times (ms) for numerical and physical comparisons of Arabic digits with control subjects (Experiment 1)

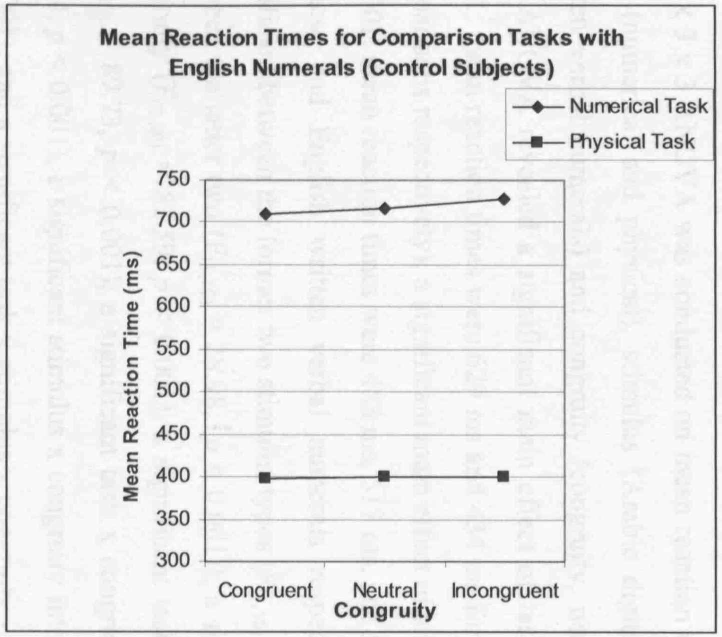


Figure 2.7 Mean reaction times (ms) for numerical and physical comparisons of English numerals with control subjects (Experiment 1)

2.3.2.2.2 Bilingual Subjects

A 2 x 3 x 3 ANOVA was conducted on mean reaction times. The factors were task (numerical and physical), stimulus (Arabic digits, Chinese and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of task ($F_{(1,10)} = 105.42$, $p < 0.001$; mean reaction times were 629 ms and 434 ms for numerical and physical comparisons respectively); a significant main effect of stimulus ($F_{(2,20)} = 35.55$, $p < 0.001$; mean reaction times were 488 ms, 517 ms, and 591 ms for Arabic digits, Chinese and English written verbal numerals respectively (difference was significant between the former two stimulus types ($F_{(1,10)} = 13.15$, $p < 0.050$) and between the latter two ($F_{(1,10)} = 28.98$, $p < 0.001$)), a significant main effect of congruity ($F_{(2,20)} = 82.59$, $p < 0.001$), a significant task x stimulus interaction ($F_{(2,20)} = 89.73$, $p < 0.001$), a significant task x congruity interaction ($F_{(2,20)} = 21.84$, $p < 0.001$), a significant stimulus x congruity interaction ($F_{(4,40)} = 8.50$, $p < 0.001$), and a significant task x stimulus x congruity interaction ($F_{(4,40)} = 2.09$, $p < 0.050$).

One-way ANOVAs were performed on mean reaction times at each level of stimulus. For Arabic digits, physical comparisons were significantly faster than numerical comparisons ($F_{(1,10)} = 83.56$, $p < 0.001$; mean reaction times were 433 ms and 544 ms respectively). For Chinese written verbal numerals, physical comparisons were significantly faster than numerical comparisons ($F_{(1,10)} = 29.44$, $p < 0.001$; mean reaction times were 457 ms and 576 ms respectively). For English written verbal numerals, physical comparisons were significantly faster than numerical comparisons ($F_{(1,10)} = 142.39$, $p < 0.001$; mean reaction times were 413 ms and 769 ms respectively).

Considering only the numerical task, a 3 x 3 ANOVA was conducted on mean reaction times. The factors were stimulus (Arabic digits, Chinese and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of stimulus ($F_{(2,20)} = 63.42$, $p <$

0.001), a significant main effect of congruity ($F_{(2,20)} = 50.46$, $p < 0.001$), and a significant stimulus x congruity interaction ($F_{(2,23)} = 5.94$, $p < 0.050$).

Further analyses revealed that Arabic digits were responded to significantly faster than Chinese written verbal numerals ($t_{(10)} = 4.95$, $p \leq 0.050$), which were in turn responded to faster than English written verbal numeral ($t_{(10)} = 64.04$, $p < 0.001$); mean reaction times were 544 ms, 576 ms, and 769 ms respectively. For Arabic digits, congruent trials were responded to significantly faster than neutral trials ($F_{(1,10)} = 20.24$, $p \leq 0.001$), which were in turn responded to significantly faster than incongruent trials ($F_{(1,10)} = 15.69$, $p < 0.050$). The mean reaction times were 511 ms, 547 ms, and 575 ms respectively (see Figure 2.8). For Chinese written verbal numerals, the same pattern emerged; congruent trials were responded to significantly faster than neutral trials ($F_{(1,10)} = 98.65$, $p < 0.001$), which were in turn responded to significantly faster than incongruent trials ($F_{(1,10)} = 54.05$, $p < 0.001$). The mean reaction times were 536 ms, 565 ms, and 627 ms respectively (see Figure 2.9). On the other hand, for English written verbal numerals, mean reaction time for congruent trials did not differ significantly from that for neutral trials ($F_{(1,10)} < 1$, *n.s.*), but the latter were responded to significantly to faster than incongruent trials ($F_{(1,10)} = 7.65$, $p < 0.050$). The mean reaction times were 757 ms, 756 ms, and 793 ms (see Figure 2.10).

Considering only the physical task, a 3 x 3 ANOVA was conducted on mean reaction times. The factors were stimulus (Arabic digits, Chinese and English written verbal numerals) and congruity (congruity, neutral, and incongruent). The ANOVA revealed a significant main effect of stimulus ($F_{(2,20)} = 13.37$, $p < 0.001$), a significant main effect of congruity ($F_{(2,20)} = 17.14$, $p < 0.001$), and a marginally significant stimulus x congruity interaction ($F_{(4,40)} = 2.55$, $p \leq 0.054$).

Further analyses revealed that Chinese written verbal numerals were responded to significantly slower than Arabic digits ($t_{(10)} = 12.84$, $p < 0.050$), which were in turn responded to significantly slower than English written verbal numerals ($t_{(10)} = 28.85$, $p < 0.001$); mean reaction times were 457 ms, 433 ms, and 413 ms respectively. For Arabic digits, mean reaction time for congruent trials did not differ significantly from that for neutral trials ($F_{(1,10)} < 1$, *n.s.*), but the latter were responded to significantly faster than incongruent trials ($F_{(1,10)} = 14.12$, $p < 0.050$). The mean reaction times were 426 ms, 426 ms, and 446 ms respectively (see Figure 2.8). For Chinese written verbal numerals, the same pattern emerged; mean reaction time for congruent trials did not differ significantly from that for neutral trials ($F_{(1,10)} < 1$, *n.s.*), but the latter were responded to significantly faster than incongruent trials ($F_{(1,10)} = 6.85$, $p < 0.050$). The mean reaction times were 451 ms, 453 ms, and 468 ms respectively (see Figure 2.9). On the other hand, English written verbal numerals did not show any significant difference in mean reaction time across the congruity levels (all *n.s.*). The mean reaction times were 412 ms, 411 ms, and 415 ms (see Figure 2.10).

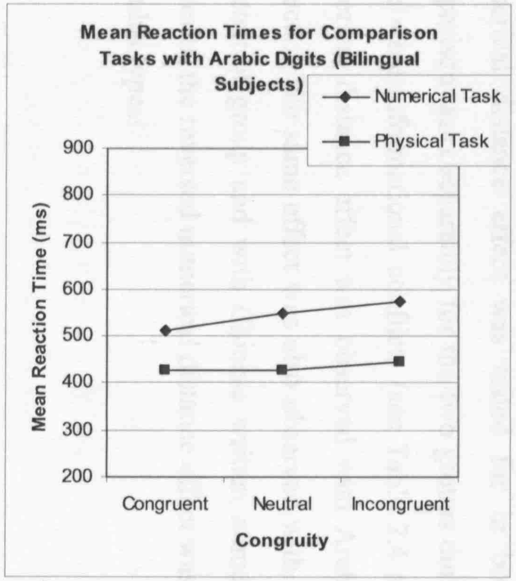


Figure 2.8 Mean reaction times (ms) for numerical and physical comparisons of Arabic digits with bilingual subjects (Experiment 1)

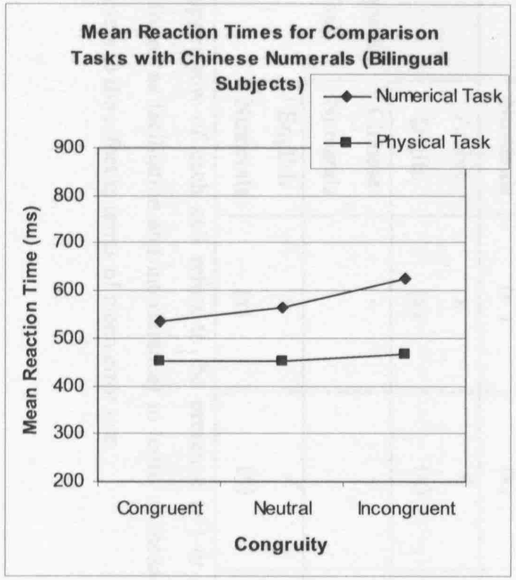


Figure 2.9 Mean reaction times (ms) for numerical and physical comparisons of Chinese numerals with bilingual subjects (Experiment 1)

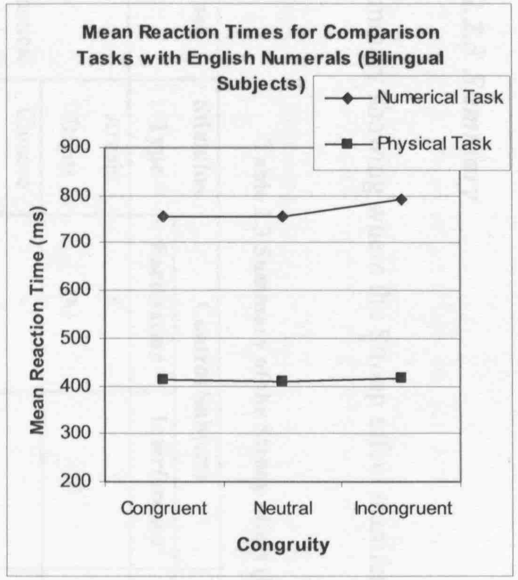


Figure 2.10 Mean reaction times (ms) for numerical and physical comparisons of English numerals with bilingual subjects (Experiment 1)

2.3.2.2.3 Summary

A summary showing where the Stroop effect manifested is given in Table 2.3.

Table 2.3 Summary of the Stroop effect (Experiment 1)

Task	Stimulus Type	Control Subjects		Bilingual Subjects	
		Facilitation	Interference	Facilitation	Interference
Numerical Task	Arabic Digits	✓ (x)	✓ (x)	✓ (✓)	✓ (x)
	Chinese Numerals	-	-	✓ (x)	✓ (x)
	English Numerals	x (✓)	x (x)	x (x)	✓ (x)
Physical Task	Arabic Digits	x (x)	✓ (x)	x (x)	✓ (x)
	Chinese Numerals	-	-	x (x)	✓ (x)
	English Numerals	x (x)	✓ (x)	x (x)	x (x)

The upper row of each cell refers to the presence (✓) or absence (x) of the Stroop effect (manifested as facilitation and interference) in terms of mean reaction time, whereas the lower row refers to the effect in terms of mean error rate.

2.3.2.2.4 Numerical Distance Effect

Numerical distance effect was tested for in both numerical and physical comparison tasks separately for the two groups during incongruent trials where there were informational conflicts (see Table 2.4 and Table 2.5). The classic numerical distance effect was observed with Arabic digits in both groups of subjects. The same effect was also observed with English written numerals in the control group and with Chinese written numerals in the bilingual group. However, the reversed numerical distance effect was not observed throughout all stimulus types.

Table 2.4 Mean reactions times (ms) for incongruent trials and t-tests comparing numerically close and far pairs at each level of stimulus type in numerical and physical comparison tasks with control subjects (Experiment 1)

Task	Stimulus Type	Close Pairs	Distance Pairs	$t_{(9)}$	Sig.
Numerical Task	Arabic Digits	606	553	6.29	$p < 0.001$
	English Numerals	744	704	4.39	$p < 0.005$
Physical Task	Arabic Digits	431	447	1.91	*n.s.*
	English Numerals	399	403	< 1	*n.s.*

Table 2.5 Mean reactions times (ms) for incongruent trials and t-tests comparing numerically close and far pairs at each level of stimulus type in numerical and physical comparison tasks with bilingual subjects (Experiment 1)

Task	Stimulus Type	Close Pairs	Distance Pairs	$t_{(10)}$	Sig.
Numerical Task	Arabic Digits	598	542	5.95	$p < 0.001$
	Chinese Numerals	657	584	6.67	$p < 0.001$
	English Numerals	798	784	1.07	*n.s.*
Physical Task	Arabic Digits	438	457	1.87	*n.s.*
	Chinese Numerals	467	469	< 1	*n.s.*
	English Numerals	415	416	< 1	*n.s.*

2.3.3 Discussion

In the present experiment, mean error rates, as well as mean reaction times, were thoroughly examined to ensure that the stimulus types were matched properly in terms of difficulty.

Error rates were generally low (mean < 6%). Control and bilingual subjects performed in a similar fashion. During numerical comparisons, the bilingual subjects made significantly more errors with English written verbal numerals than with Arabic digits. The control subjects showed a similar pattern – mean error rate was marginally higher with English written numerals than with Arabic digits. In addition, the bilingual subjects performed the task with Chinese written verbal numerals and these stimuli showed a significantly lower mean error rate relative to English written verbal numerals.

Mean error rates for physical comparisons were low (all < 1%). Unlike the pilot study, there was no evidence to suggest that Chinese written verbal numerals were less physically discriminable than the other stimulus types in the current experiment; the mean error rate for Chinese stimuli were comparable to those for Arabic and English stimuli, and mean reaction time for the former was not unexpectedly high.

In terms of mean reaction time, control and bilingual subjects performed in a similar fashion during numerical comparisons – both groups were significantly faster with Arabic digits than with English written verbal numerals. On the other hand, physical comparisons were responded to faster when the stimuli were English written verbal numerals than when they were Arabic digits for both groups of subjects. This could be explained by the larger area differences between English numerals than between Arabic digits, for example, the physical size difference between the pair (SEVEN SEVEN) is more prominent than that between (7 7), hence the former would require less time to process. It must be noted that the reaction time differences between these stimulus types were very different in scale across the two tasks – the difference during numerical comparisons was much larger than that during physical comparisons.

For the bilingual subjects, Chinese written verbal numerals were responded to significantly slower than Arabic digits, but significantly faster than English written verbal numerals during numerical comparisons. During physical comparisons, Chinese numerals were responded to significantly slower than the other two stimulus types. This might be due to the large variation in maximum heights of Chinese written verbal numerals, making them harder to discriminate physically. It must be noted that even though there were significant differences across stimulus types during physical comparisons, all the mean reaction times were fast (all < 460 ms). Thus, there was no reason to believe that any Stroop effect observed with Chinese written verbal numerals had been exaggerated by the difficulty arisen from the varying heights of the stimuli.

2.3.3.1 The Stroop Effect and the Effects of Writing System

Since error rates for the present experiment were generally low, this measure could not be considered a sensitive indicator for the Stroop effect (see Table 2.3). Consequently, the rest of the discussion will focus on mean reaction time analyses.

In terms of mean reaction time, the Stroop effect (manifested as facilitation and interference) was observed with both groups of subjects with Arabic digits during numerical comparisons. During physical comparisons, both groups exhibited an interference effect with Arabic digits. Moreover, the control subjects showed an interference effect with English written verbal numerals, whereas the bilingual subjects showed the same effect with Chinese written verbal numerals. During numerical comparisons, the bilingual subjects exhibited the Stroop effect (manifested as both facilitation and interference) with Chinese written verbal numerals, but experienced only the interference effect with English written verbal numerals.

The findings have provided solid evidence the Stroop effect was bi-directional; this was most clearly shown when considering comparisons with Arabic digits. For all stimulus types, mean reaction time was significantly faster for physical comparisons than for numerical comparisons. The present experiment replicated previous research findings (Girelli et al., 2000; Henik & Tzelgov, 1982) that for Arabic digits, (1) the mean reaction time for the physical comparison task was significantly faster than that for the numerical comparison task, and (2) the Stroop effect was stronger in the numerical task (both interference and facilitation were observed) than in the physical task (only interference was observed). The findings provide support for Foltz et al.'s (1984) suggestion that "physical size as an irrelevant dimension had more influence on performance than did semantic size as an irrelevant dimension."

The present findings show that mean reaction time for Arabic digits was significantly faster than that for English written verbal numerals in both groups of subjects. In addition, the mean reaction time for Chinese written verbal numerals was significantly faster than that for English written verbal numerals and significantly slower than that for Arabic digits in the bilingual group. Furthermore, for Arabic digits and Chinese written verbal numerals, the two comparison tasks showed a smaller difference in terms of mean reaction time (< 140 ms), whereas the difference in mean reaction time between the tasks for English written verbal numerals was relatively larger (> 310 ms). According to Foltz et al.'s (1984) relative speed of processing account, the Stroop effect would occur when the two comparison tasks occur concurrently and it would be reasonable to assume that the more similar the mean reaction times between the tasks, the more overlap there would be between the two comparison processes. Hence, Foltz et al.'s (1984) account would predict a stronger Stroop effect with Arabic digits and Chinese written verbal numerals since they appeared to show more processing overlap compared to English written verbal numerals. Moreover, this account would also predict an inverse relation between processing time and the severity of the Stroop effect, i.e., the faster a process, the more influence it would have on the slower process, thus predicting a strong Stroop effect with Arabic digits, and a weaker one with English written verbal numerals. The current findings with the bilingual subjects appear to support such a prediction; for the numerical task, both facilitation and inference were observed in the numerical task with Arabic digits and Chinese written verbal numerals, but only (weak) interference was observed with English written verbal numerals, and for the physical task, the former two stimulus types showed an interference effect, whereas the latter did not show any Stroop effect.

However, when considering both groups of subjects, results with English written verbal numbers could not be explained by Foltz et al.'s (1984) relative speed of processing. Whilst the bilingual subjects showed an interference effect in the numerical task, the control subjects exhibited the same effect in the physical task. Foltz et al.'s (1984) predicted a stronger Stroop effect in the numerical task than in the physical task, thus failing to account for this discrepancy.

On the other hand, Cohen et al.'s (1990) parallel distributed processing model, with emphasis placed on the relative strength of process, provides a better account for the findings. Despite the efforts to select bilingual subjects who were competent in English, control subjects – being native English speakers – were almost certainly more proficient in the English language than the bilingual subjects to whom English was their second language. Thus, according to Cohen et al. (1990), the control subjects would have a stronger semantic pathway in English compared to the bilingual subjects. Consequently, it would be reasonable to expect numerical magnitudes to have a stronger influence on the physical comparisons in the control subjects compared to the bilingual subjects. Following the same argument, the interference effect observed in the bilingual subjects during numerical comparisons could be explained by a weaker English semantic pathway, thus more susceptible to influences from the physical dimension.

The finding that interference was observed during physical comparison with both Chinese and English written verbal numerals (in bilingual and control subjects respectively) is consistent with Yin and Butterworth's (1992) suggestion that reading processes involved in logographic and alphabetic writing systems have the same broad underlying cognitive architecture.

The Stroop effect observed with Chinese but not English written verbal numerals during physical comparisons in bilingual subjects is consistent with the earlier suggestion that the direct orthography-to-semantics route is stronger and preferred compared to the indirect route via phonology when processing Chinese stimulus, and according to Cohen et al. (1990), the stronger the pathway of a process is, the less attention it requires, and the more interference it induces on the task-relevant dimension. The strength of the orthography-to-semantics route in Chinese may have arisen from the salient semantic feature of visual character form. This is true for at least the small numbers 一 (one), 二 (two), and 三 (three) which are very self-evident in Chinese. In contrast, English written verbal numerals whose visual forms do not possess such a salient semantic feature, are less capable of inducing a Stroop effect as that produced by Chinese written verbal numerals.

It is important to reiterate that the relatively lower proficiency with English in the bilingual subjects might also have contributed to the absence of an interference effect with English written verbal numerals during physical comparisons. This explanation is consistent with Girelli et al.'s (2000) finding that young children who lacked a comprehensive understanding of number semantics showed no Stroop effect during physical comparisons.

2.3.3.2 Numerical Distance Effect

The classic numerical distance effect was observed during numerical comparisons with Arabic digits in both groups of subjects, with Chinese written verbal numerals in the bilingual group, and with English written verbal numerals in the control group. The findings suggest that when numerical magnitude was the task-relevant dimension, refined processing took place. The absence of the numerical distance effect with English written verbal numerals, once again, suggests that the bilingual subjects were less proficient in English than the control subjects.

When numerical magnitude was the task-irrelevant dimension, i.e., during physical comparisons, the reversed numerical distance effect, reported by Henik and Tzelgov (1982) and Girelli et al. (2000), failed to replicate. It is however important to note that the differences in mean reaction time were all in the predicted direction, i.e., in the direction of a reversed numerical distance effect. As suggested in Chapter 1, the use of a factorial design, as in the present experiment, might have limited inferences that could be implicated on the amount of information available in the task-relevant and -irrelevant dimensions. The absence of the reversed numerical distance effect might have resulted from the use of a factorial design – using just two levels of numerical distance might not have been enough for the reversed numerical distance effect to emerge. By using only two levels, only a mere difference between numerically close and distant conditions could be established, but not a linear change (decrease/ increase in mean reaction time) with numerical distance. Given these problems, it would be hasty to jump to the conclusion that numerical magnitudes are not processed in a refined fashion during task-irrelevant conditions based on the absence of a reversed numerical distance effect.

The present findings also highlight the limitations in inferences which could be made from the Stroop effect. For instance, interference was observed with Arabic digits during the physical comparison task in the control subjects. If this finding was interpreted as support for autonomous processing, then the absence of a reversed numerical distance effect would suggest that autonomous processing of numerical magnitudes was coarse (i.e., merely a large-small dichotomous classification rather refined judgements on a graded scale). However, others have reported the latter effect (e.g., Girelli et al., 2000; Henik and Tzelgov, 1982), suggesting that autonomous processing of numerical magnitudes under task-irrelevant condition was refined. Thus, findings point to the need of a better experimental design. Until a fully parametric design is put to test, the precise nature of numerical magnitude processing under task-irrelevant conditions cannot be verified.

2.4 Experiment 2: Comparing Numerical Magnitudes in Japanese and in English

Experiment 2 was designed to investigate the processing of numerical information in the two Japanese scripts, Kanji and Kana. Japanese numerals are commonly written in Kanji, an ideographic script, but rarely in Kana, a syllabic script. The present experiment examined the similarities and differences in processing numerical magnitudes written in these scripts, compared to Arabic digits and numerals written in English – a language which the subjects only learned later in life.

2.4.1 Methods

2.4.1.1 Tasks

The two tasks were numerical magnitude comparison and physical size comparison tasks. Subjects had to select, via a key response (by pressing the left "F" key with their left index finger or the right "J" key with their right index on a qwerty keyboard) the larger number of the pair in numerical magnitude or in physical size accordingly. Their reaction times and responses were recorded.

The tasks were computer-based. Two programs were used: one written in Cogent (running on a MATLAB Version 6.1 platform) and other written in JavaScript. These were included so that program equivalence could be examined.

2.4.1.2 Stimuli

In each trial, two numbers appeared in white on a black background in the middle of a 15" TFT screen. Each presentation lasted 1000 ms, and there was an interval of 1000 ms before the subsequent trial. The digits were approximately 14.50 cm apart, measured between the centres of the two stimuli. Subjects sat at arm's length from the screen.

There were 4 types of stimulus: Arabic digits, Kanji written verbal numerals and their Kana equivalents, and English written verbal numerals. The numbers used were 1, 3, 4, 6, 7, and 8; these were chosen, so that in the Kana conditions, only number equivalents with two syllables would be displayed.

6 levels of physical size (1, 3, 4, 6, 7, and 8) were used; they were created with constant ratio (in height) of 1.445 between adjacent ones (thus, between 1 and 3, the ratio was 1.445^2). The physical sizes of the Arabic digits appeared on the screen were approximately 0.6 cm, 1.1 cm (unit 1), 1.8 cm (unit 3), 3.1 cm (unit

4), 5.1 cm (unit 6), and 7.4 cm (unit 8) in height. The dimensions of the other stimuli varied, but were created with the same constant ratio.

There were 3 experimental conditions: congruent, when the numerically larger digit was physically larger (e.g., 3 6); incongruent, when the numerically larger digit was physically smaller (e.g., 3 6); and neutral, when the two digits were of the same physical size in the numerical comparison task (e.g., 3 6) or when the same digit appeared in different sizes in the physical comparison task (e.g., 3 3).

Three levels of numerical distance (1, 2, and 3) and three levels of physical distance were used. The pairs used were: 3 4, 6 7, 7 8 (for numerical distance of 1); 1 3, 4 6, 6 8 (for numerical distance of 2); 1 4, 3 6, 4 7 (for numerical distance of 3). Each of these pairs were presented twice in physical sizes belonging to each of the 3 levels of physical distance, once as congruent and once as incongruent trial, making a total of 27 trials of each type. The equivalent pairs were used in the physical comparison task, e.g., for physical size of 1 unit, the pair would consist of a number in size unit 3 (1.1 cm in height for digit) and another in size unit 4 (1.8 cm in height for digit). In the numerical comparison task, the neutral trials consisted of the same 9 pairs listed above, each appearing 3 times in different physical sizes, making a total of 27 neutral trials. In the physical comparison task, the pairs used were: 1 1, 3 3, 4 4, 6 6, 7 7, 8 8; each pair appeared once or twice in physical sizes belonging to each of the 3 physical distance groups, making a total of 27 neutral trials.

The total number of trials in each task was 81, approximately half of which had the correct response appearing on the left, and the rest on the right. Stimuli were presented in a pseudorandom order with the following constraints to avoid carryover effects: (1a) the same digit did not appear on the same side in consecutive trials in the numerical comparison task, (1b) the same physical size did not appear on the same side in consecutive trials in the physical comparison task, (2) the correct response did not appear on the same side (left or right) for more than three consecutive trials, and (3) the experimental condition (i.e., congruent, incongruent, and neutral) was not the same for more than two consecutive trials.

2.4.1.3 Subjects

48 Japanese subjects (24 males and 24 females), age ranged 18 to 36 (mean = 23.9 years, standard deviation = 4.5 years) took part. All the subjects had studied English as their second language at high school (which started at the age of 13) for at least 6 years. Half of the subjects were tested with the Cogent program, and half with the JavaScript program. They all performed 8 conditions (4 numerical comparison tasks and 4 physical comparison tasks). The comparison tasks alternated: half of subjects started with a numerical task, while the other half a physical task. The order of stimulus type was randomized.

2.4.2 Results

ANOVAs were used to analyse the mean percentage errors and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

2.4.2.1 Analyses on Mean Error Rates

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 1000 ms after the stimulus offset.

A 2 x 4 x 2 mixed design ANOVA was conducted on mean error rates. The repeated factors were task (numerical and physical comparison tasks) and stimulus type (digits, Kanji, Kana, and English), and the between factor was program (Cogent and JavaScript). The ANOVA revealed a significant main effect of task ($F_{(1,46)} = 77.84$, $p < 0.001$), a significant main effect of stimulus type ($F_{(2,107)} = 35.38$, $p < 0.001$), and a significant task x stimulus type interaction ($F_{(2,93)} = 40.88$, $p < 0.001$). The main effect of program was non-significant ($F_{(1,46)} = 1.30$, *n.s.*) and this factor did not significantly interact with any other factor (all *n.s.*).

After collapsing mean error rates over the factor program, the mean error rate for numerical comparisons was 7.60% and that for physical comparisons was 1.36%. Figure 2.11 shows the mean error rates for each stimulus type at each level of task.

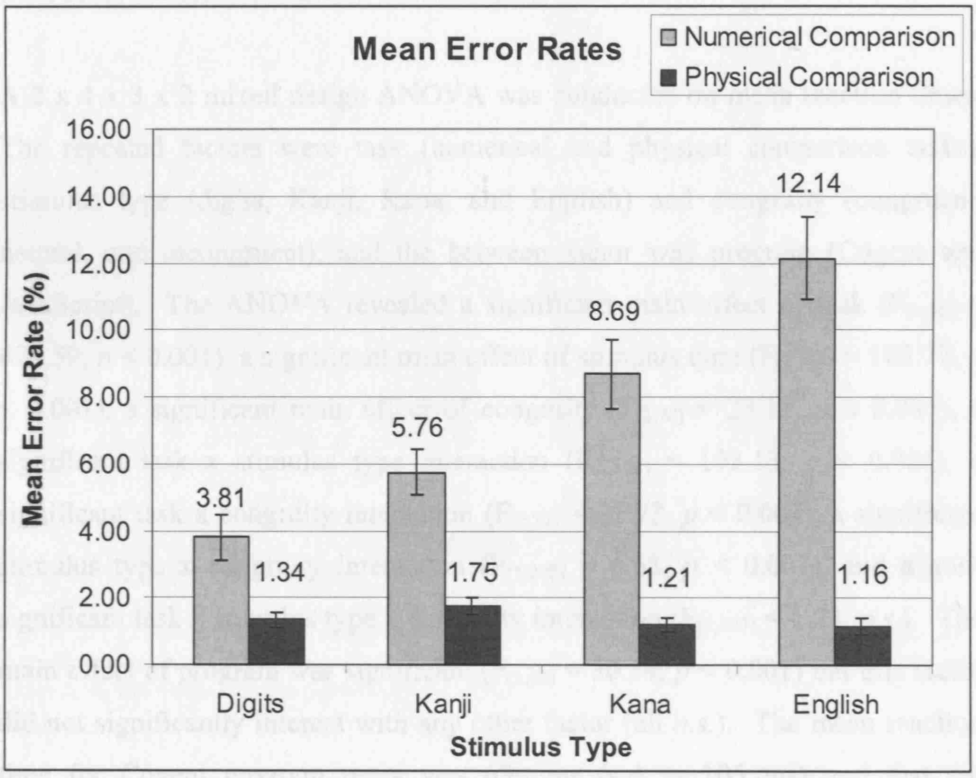


Figure 2.11 Mean error rates for numerical and physical comparison tasks with Arabic digits, Kanji written verbal numerals and their Kana equivalents, and English written verbal numerals (Experiment 2)

A one-way ANOVA was conducted at each level of task. For the numerical comparison task, the ANOVA revealed a significant main effect of stimulus type ($F_{(2,98)} = 42.68$, $p < 0.001$). Tests of within-subjects contrasts revealed significant differences in mean error rates between Arabic digits and Kanji number words ($F_{(1,47)} = 18.52$, $p < 0.001$), between Kanji written verbal numerals and Kana number equivalents ($F_{(1,47)} = 18.30$, $p < 0.001$), and between Kana number equivalents and English written verbal numerals ($F_{(1,47)} = 19.82$, $p <$

0.001). For the physical comparison task, the ANOVA revealed a non-significant main effect of stimulus type ($F_{(3,141)} = 2.22$, *n.s.*).

2.4.2.2 Analyses on Mean Reaction Times

Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were removed prior to reaction time analyses.

A 2 x 4 x 3 x 2 mixed design ANOVA was conducted on mean reaction times. The repeated factors were task (numerical and physical comparison tasks), stimulus type (digits, Kanji, Kana, and English) and congruity (congruent, neutral, and incongruent), and the between factor was program (Cogent and JavaScript). The ANOVA revealed a significant main effect of task ($F_{(1,46)} = 678.59$, $p < 0.001$), a significant main effect of stimulus type ($F_{(2,112)} = 187.77$, $p < 0.001$), a significant main effect of congruity ($F_{(2,70)} = 23.18$, $p < 0.001$), a significant task x stimulus type interaction ($F_{(2,107)} = 193.13$, $p < 0.001$), a significant task x congruity interaction ($F_{(2,77)} = 21.73$, $p < 0.001$), a significant stimulus type x congruity interaction ($F_{(5,220)} = 6.68$, $p < 0.001$), and a non-significant task x stimulus type x congruity interaction ($F_{(5,207)} = 1.23$, *n.s.*). The main effect of program was significant ($F_{(1,46)} = 30.54$, $p < 0.001$) but this factor did not significantly interact with any other factor (all *n.s.*). The mean reaction time for Cogent program users was 690 ms (s.d. = 105 ms) and that for JavaScript program users was 571 ms (s.d. = 84 ms).

The factor program did not interact with any other factor suggesting that the patterns of reaction time for the different comparison tasks did not differ between the groups who were tested with different programs. Therefore further analyses were conducted with the mean reaction times collapsed over this factor.

After collapsing mean reaction times over the factor program, the mean reaction time for numerical comparisons was 711 ms and that for physical comparisons was 491 ms. Figure 2.12 shows the mean reaction times for each stimulus type at each level of task.

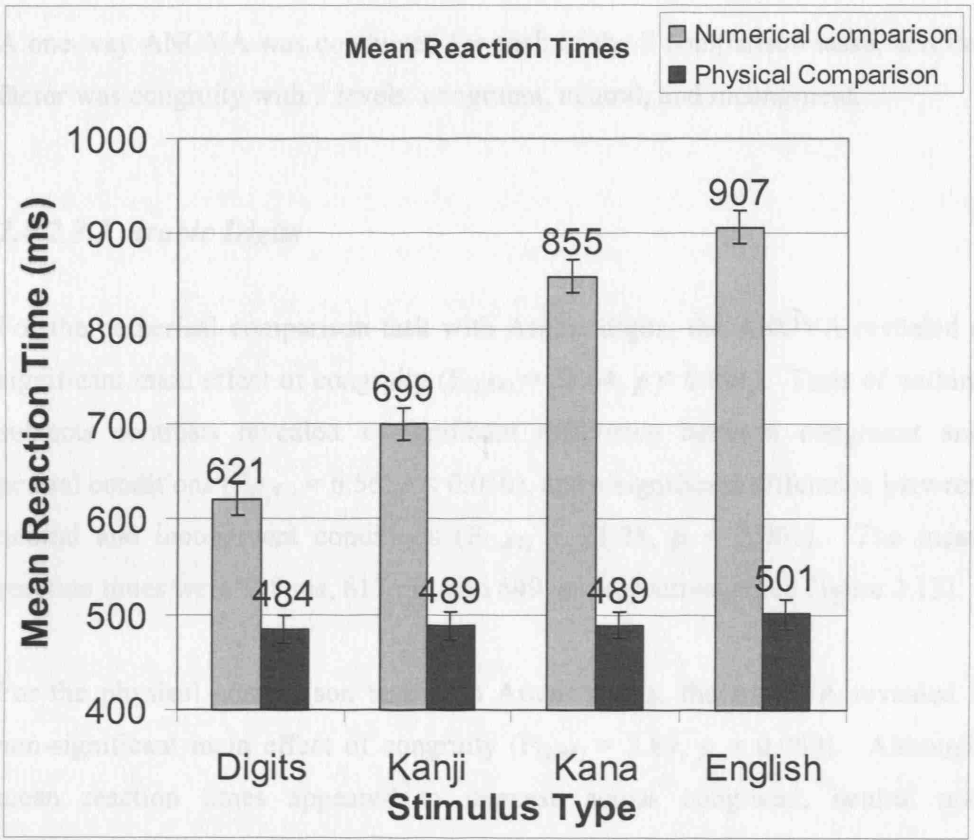


Figure 2.12 Mean reaction times (ms) for numerical and physical comparison tasks with Arabic digits, Kanji written verbal numerals and their Kana equivalents, and English written verbal numerals (Experiment 2)

A one-way ANOVA was conducted at each level of task to test for differences between the four stimulus types. For the numerical comparison task, the ANOVA revealed a significant main effect of stimulus type ($F_{(2,105)} = 244.53, p < 0.001$). Tests of within-subjects contrasts revealed significant differences in mean reaction times between Arabic digits and Kanji written verbal numerals ($F_{(1,47)} = 85.99, p < 0.001$), between Kanji written verbal numerals and Kana number equivalents ($F_{(1,47)} = 206.87, p < 0.001$), and between Kana number equivalents and English written verbal numerals ($F_{(1,47)} = 21.24, p < 0.001$). For the physical comparison task, the ANOVA revealed a non-significant main effect of stimulus type ($F_{(3,141)} = 2.57$, *n.s.*).

2.4.2.3 Testing for the Stroop Effect

A one-way ANOVA was conducted for each of the 8 comparison tasks, and the factor was congruity with 3 levels: congruent, neutral, and incongruent.

2.4.2.3.1 Arabic Digits

For the numerical comparison task with Arabic digits, the ANOVA revealed a significant main effect of congruity ($F_{(2,94)} = 23.64$, $p < 0.001$). Tests of within-subjects contrasts revealed a significant difference between congruent and neutral conditions ($F_{(1,47)} = 6.56$, $p < 0.050$), and a significant difference between neutral and incongruent conditions ($F_{(1,47)} = 21.28$, $p < 0.001$). The mean reaction times were 598 ms, 617 ms, and 649 ms respectively (see Figure 2.13).

For the physical comparison task with Arabic digits, the ANOVA revealed a non-significant main effect of congruity ($F_{(2,94)} = 2.89$, $p = 0.060$). Although mean reaction times appeared to increase across congruent, neutral and incongruent (476 ms, 485 ms, 491 ms respectively), the main effect of congruity failed to reach significance (see Figure 2.13).

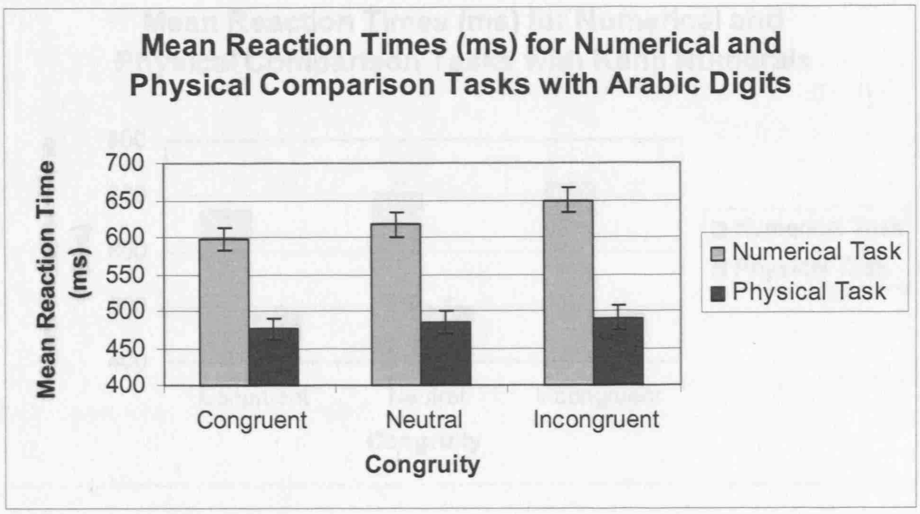


Figure 2.13 Mean reaction times (ms) for numerical and physical comparison tasks with Arabic digits (Experiment 2)

2.4.2.3.2 Kanji Written Verbal Numerals

For the numerical comparison task with Kanji written verbal numerals, the ANOVA revealed a significant main effect of congruity ($F_{(1,64)} = 13.29$, $p < 0.001$). Tests of within-subjects contrasts revealed a significant difference between congruent and neutral conditions ($F_{(1,47)} = 25.39$, $p < 0.001$), and a non-significant difference between neutral and incongruent conditions ($F_{(1,47)} = 3.74$, $p = 0.059$). The mean reaction times were 675 ms, 703 ms, and 720 ms respectively (see Figure 2.14).

For the physical comparison task with Kanji written verbal numerals, the ANOVA revealed a significant main effect of congruity ($F_{(2,94)} = 4.64$, $p < 0.050$). Tests of within-subjects contrasts revealed a significant difference between congruent and neutral conditions ($F_{(1,47)} = 6.42$, $p < 0.050$), and a significant difference between neutral and incongruent conditions ($F_{(1,47)} = 7.34$, $p < 0.010$). The mean reaction times were 484 ms, 499 ms, and 484 ms respectively (see Figure 2.14). Note that the mean reaction time for incongruent trials was, contrary to the prediction of an interference effect, significantly faster than that for neutral trials.

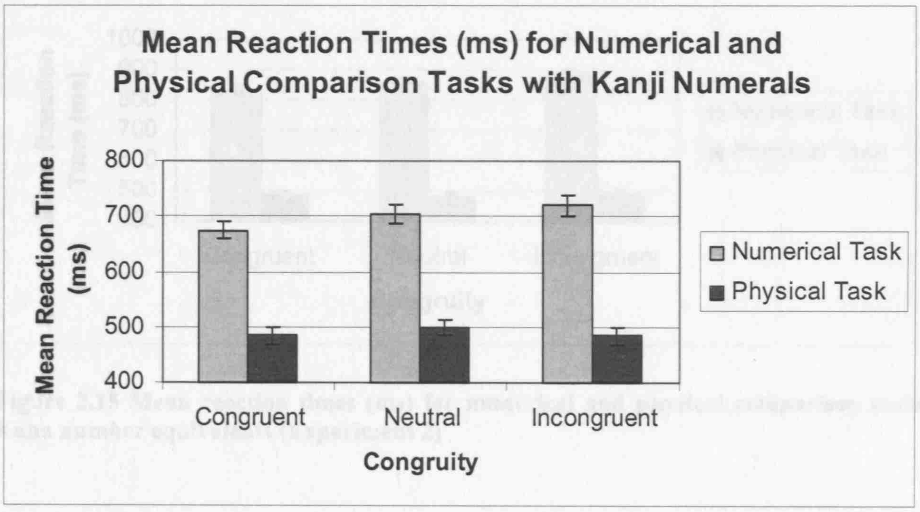


Figure 2.14 Mean reaction times (ms) for numerical and physical comparison tasks with Kanji numerals (Experiment 2)

2.4.2.3.3 Kana Number Equivalents

For the numerical comparison task with Kana number equivalents, the ANOVA revealed a significant main effect of congruity ($F_{(2,94)} = 18.24, p < 0.001$). Tests of within-subjects contrasts revealed a significant difference between congruent and neutral conditions ($F_{(1,47)} = 12.26, p \leq 0.001$), and a significant difference between neutral and incongruent conditions ($F_{(1,47)} = 8.09, p < 0.010$). The mean reaction times were 830 ms, 855 ms, and 880 ms respectively (see Figure 2.15).

For the physical comparison task with Kana number equivalents, the ANOVA revealed a significant main effect of congruity ($F_{(2,94)} = 5.63, p \leq 0.005$). Tests of within-subjects contrasts revealed a non-significant difference between congruent and neutral conditions ($F_{(1,47)} < 1$, *n.s.*), and a significant difference between neutral and incongruent conditions ($F_{(1,47)} = 10.10, p < 0.005$). The mean reaction times were 486 ms, 483 ms, and 499 ms respectively (see Figure 2.15).

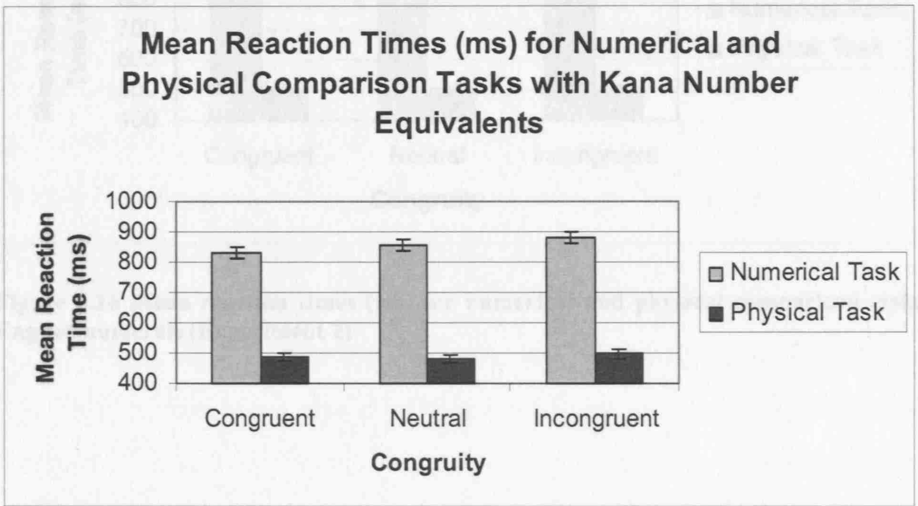


Figure 2.15 Mean reaction times (ms) for numerical and physical comparison tasks with Kana number equivalents (Experiment 2)

2.4.2.3.4 English Written Verbal Numerals

For the numerical comparison task with English written verbal numerals, the ANOVA revealed a non-significant main effect of congruity ($F_{(2,94)} = 1.33$, *n.s.*). The mean reaction times were 900 ms, 909 ms, and 911 ms for congruent, neutral, and incongruent trials respectively (see Figure 2.16).

For the physical comparison task with Arabic digits, the ANOVA revealed a non-significant main effect of congruity ($F_{(2,94)} = 1.53$, *n.s.*). The mean reaction times were 504 ms, 504 ms, and 495 ms for congruent, neutral, and incongruent trials respectively (see Figure 2.16).

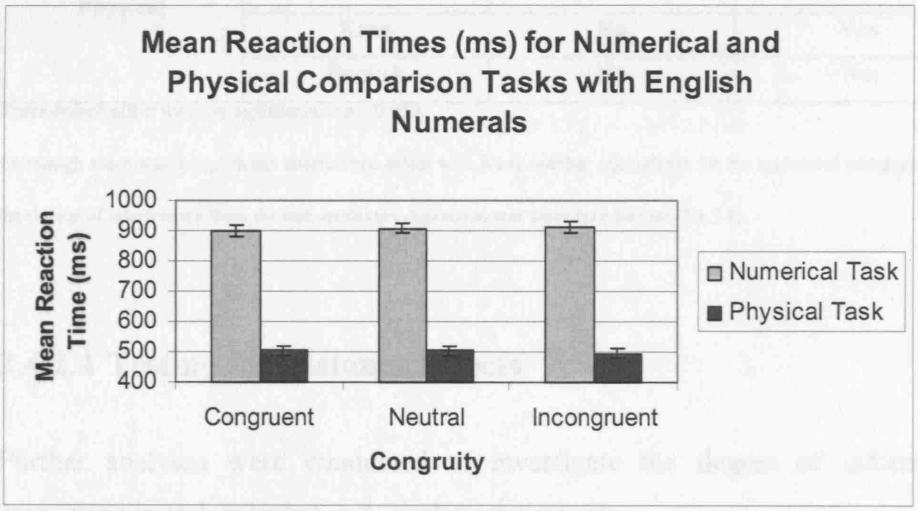


Figure 2.16 Mean reaction times (ms) for numerical and physical comparison tasks with English numerals (Experiment 2)

2.4.2.3.5 Summary

A summary showing where the Stroop effect manifested is given in Table 2.6.

Table 2.6 Summary of the Stroop effect (manifested as facilitation and interference) across different stimulus types (Experiment 2)

Comparison Task	Stimulus Type	Facilitation	Interference
Numerical	Digit	Yes	Yes
	Kanji	Yes	No#
	Kana	Yes	Yes†
	English	No	No
Physical	Digit	No	No
	Kanji	Yes	No
	Kana	No	Yes
	English	No	No

#Interference effect was non-significant at $p = 0.059$.

†Although there was a significant interference effect with Kana number equivalents for the numerical comparison task, the degree of interference from the task-irrelevant dimension was weak (see Section 2.4.3.4).

2.4.2.4 Testing for Distance Effects

Further analyses were conducted to investigate the degree of information processing in task-relevant and -irrelevant channels.

A 2 x 3 repeated-measures ANOVA was carried out for each comparison task. The two factors were task-relevance (task-relevant and task-irrelevant) and distance (1, 2, and 3). In each of these analyses, only mean reaction times from the incongruent conditions, where there were informational conflicts, were used.

2.4.2.4.1 Arabic Digits

For numerical comparison with Arabic digits, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,46)} < 1$, *n.s.*), a significant main effect of distance ($F_{(2,92)} = 6.80$, $p < 0.005$), and a significant task-relevance x distance interaction ($F_{(2,92)} = 13.81$, $p < 0.001$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was significant at the linear level ($F_{(1,46)} = 31.99$, $p < 0.001$), but non-significant at the quadratic level ($F_{(1,46)} < 1$, *n.s.*). A one-way ANOVA was conducted at each level of task-relevance. Numerical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,92)} = 15.40$, $p < 0.001$) and test of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,46)} = 30.34$, $p < 0.001$) for this dimension. The task-irrelevant dimension, physical distance, showed a significant main effect of distance ($F_{(2,92)} = 4.83$, $p \leq 0.010$), and test of within-subjects contrasts revealed a significant positive linear trend ($F_{(1,46)} = 9.20$, $p < 0.005$) for this dimension. Figure 2.17 shows the reaction time patterns for numerical comparison task with Arabic digits.

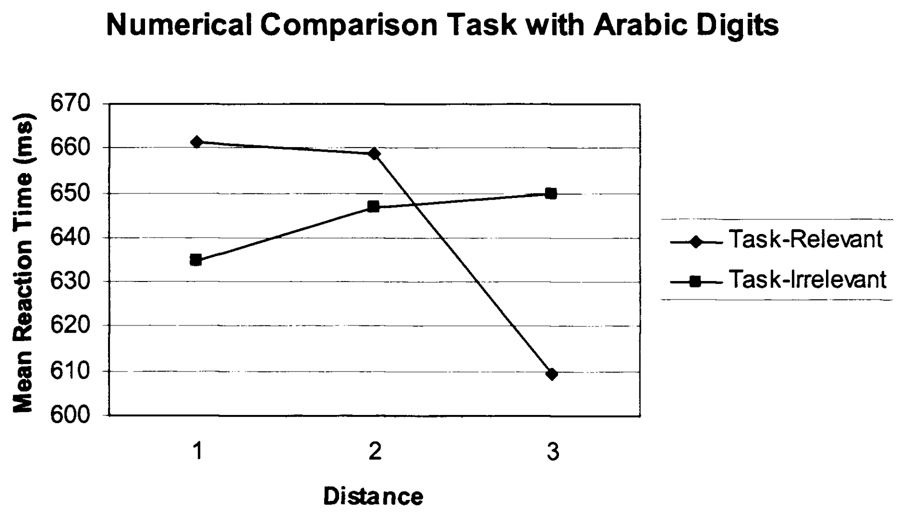


Figure 2.17 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during numerical comparisons with Arabic digits (Experiment 2)

For physical comparison with Arabic digits, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,47)} < 1$, *n.s.*), a non-significant main effect of distance ($F_{(2,82)} < 1$, *n.s.*), and a significant task-relevance x distance interaction ($F_{(2,76)} = 31.05, p < 0.001$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was significant at the linear level ($F_{(1,47)} = 38.73, p < 0.001$) and at the quadratic level ($F_{(1,47)} = 10.83, p \leq 0.002$). A one-way ANOVA was conducted at each level of task-relevance. Physical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,94)} = 20.13$, $p < 0.001$) and test of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,47)} = 31.51$, $p < 0.001$) and a significant quadratic trend ($F_{(1,47)} = 4.91, p < 0.050$) for this dimension. The task-irrelevant dimension, numerical distance, showed a significant main effect of distance ($F_{(2,94)} = 15.86$, $p < 0.001$), and test of within-subjects contrasts revealed a significant positive linear trend ($F_{(1,47)} = 26.29, p < 0.001$) and a non-significant quadratic trend ($F_{(1,47)} = 3.09$, *n.s.*) for this dimension. Figure 2.18 shows the reaction time patterns for physical comparison task with Arabic digits.

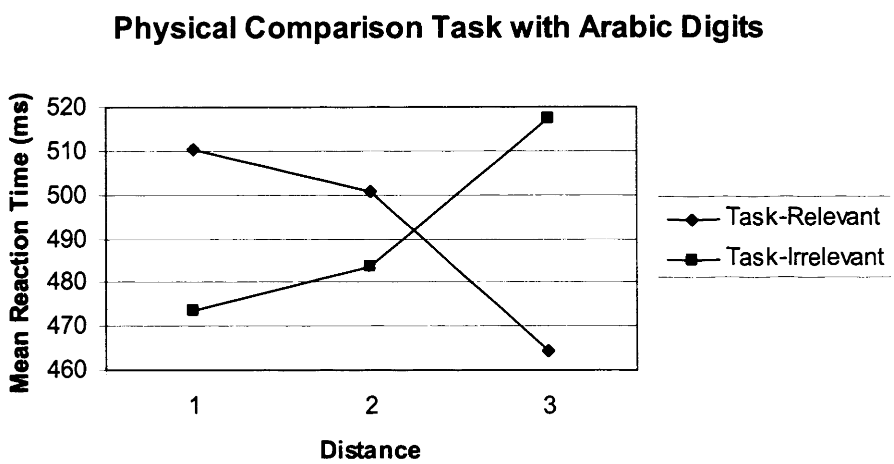


Figure 2.18 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during physical comparisons with Arabic digits (Experiment 2)

2.4.2.4.2 Kanji Written Verbal Numerals

For numerical comparison with Kanji written verbal numerals, the ANOVA revealed a significant main effect of task-relevance ($F_{(1,46)} = 11.56, p \leq 0.001$), a significant main effect of distance ($F_{(2,92)} = 18.47, p < 0.001$), and a significant task-relevance x distance interaction ($F_{(2,81)} = 78.57, p < 0.001$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was significant at the linear level ($F_{(1,46)} = 121.99, p < 0.001$) and at the quadratic level ($F_{(1,46)} = 6.08, p < 0.020$). A one-way ANOVA was conducted at each level of task-relevance. Numerical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,80)} = 72.50, p < 0.001$) and test of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,46)} = 96.25, p < 0.001$) and a significant quadratic trend ($F_{(1,46)} = 23.92, p < 0.001$) for this dimension. The task-irrelevant dimension, physical distance, showed a significant main effect of distance ($F_{(2,92)} = 10.98, p < 0.001$), and test of within-subjects contrasts revealed a significant positive linear trend ($F_{(1,46)} = 17.33, p < 0.005$) and a non-significant quadratic trend ($F_{(1,46)} = 2.25$, *n.s.*) for this dimension. Figure 2.19 shows the reaction time patterns for numerical comparison task with Kanji numerals.

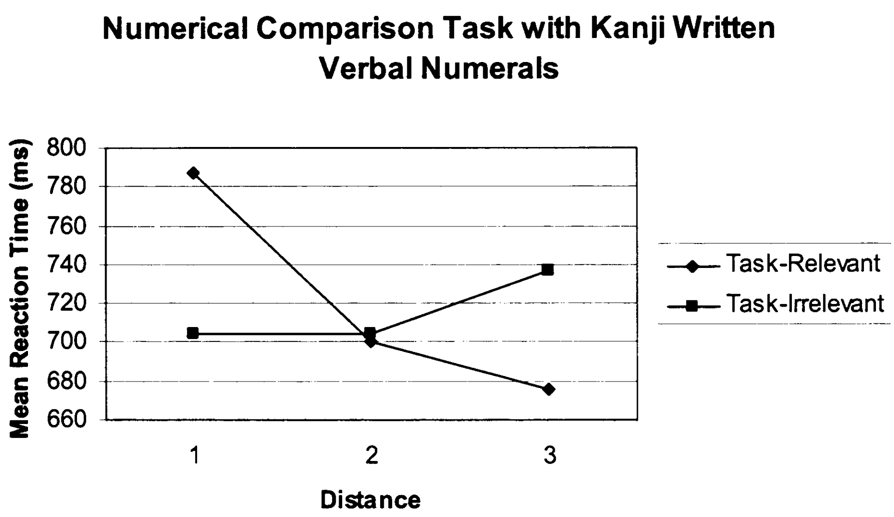


Figure 2.19 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during numerical comparisons with Kanji written verbal numerals (Experiment 2)

For physical comparison with Kanji written verbal numerals, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,47)} < 1$, *n.s.*), a significant main effect of distance ($F_{(2,77)} = 4.64$, $p < 0.020$), and a significant task-relevance x distance interaction ($F_{(2,94)} = 8.86$, $p < 0.001$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was significant at the linear level ($F_{(1,47)} = 5.44$, $p < 0.050$) and at the quadratic level ($F_{(1,47)} = 14.36$, $p < 0.001$). A one-way ANOVA was conducted at each level of task-relevance. Physical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,94)} = 4.84$, $p \leq 0.010$). Tests of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,47)} = 11.45$, $p \leq 0.001$), but a non-significant quadratic trend ($F_{(1,47)} < 1$, *n.s.*). The task-irrelevant dimension, numerical distance, showed a significant main effect ($F_{(2,75)} = 7.84$, $p \leq 0.002$). Tests of within-subjects contrasts revealed a non-significant linear trend ($F_{(1,47)} < 1$, *n.s.*), but a significant quadratic trend ($F_{(1,47)} = 19.18$, $p < 0.001$). Figure 2.20 shows the reaction time patterns for physical comparison task with Kanji numerals.

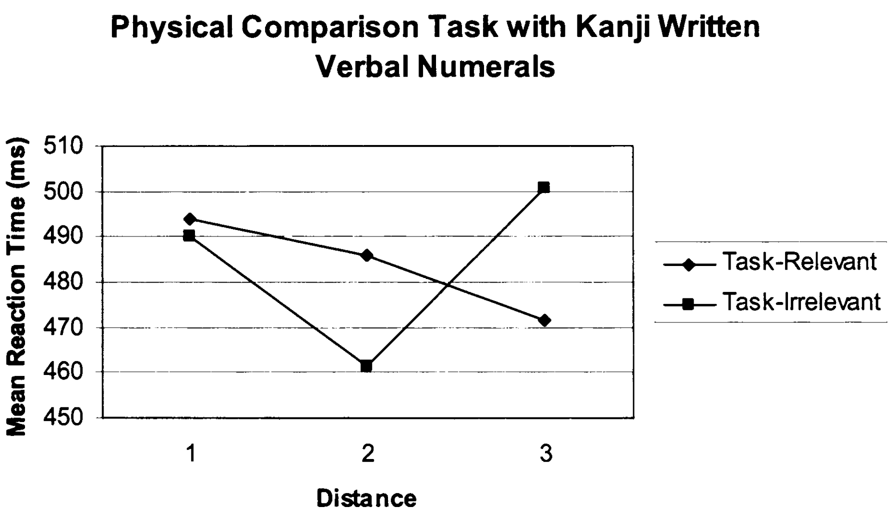


Figure 2.20 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during physical comparisons with Kanji written verbal numerals (Experiment 2)

2.4.2.4.3 Kana Number Equivalents

For numerical comparison with Kana number equivalents, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,46)} = 1.34$, *n.s.*), a significant main effect of distance ($F_{(2,80)} = 21.99$, $p < 0.001$), and a marginally significant task-relevance x distance interaction ($F_{(2,92)} = 3.03$, $p = 0.053$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was marginally significant at the linear level ($F_{(1,46)} = 3.92$, $p = 0.054$) and non-significant at the quadratic level ($F_{(1,46)} = 2.25$, *n.s.*). A one-way ANOVA was conducted at each level of task-relevance. Numerical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,92)} = 18.70$, $p < 0.001$). Tests of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,46)} = 6.02$, $p < 0.050$), and a significant quadratic trend ($F_{(1,46)} = 25.18$, $p < 0.001$). The task-irrelevant dimension, physical distance, showed a significant main effect ($F_{(2,92)} = 5.42$, $p < 0.010$). Tests of within-subjects contrasts revealed a non-significant linear trend ($F_{(1,46)} < 1$, *n.s.*), but significant quadratic trend ($F_{(1,46)} = 9.22$, $p < 0.005$). Figure 2.21 shows the reaction time patterns for numerical comparison task with Kana number equivalents.

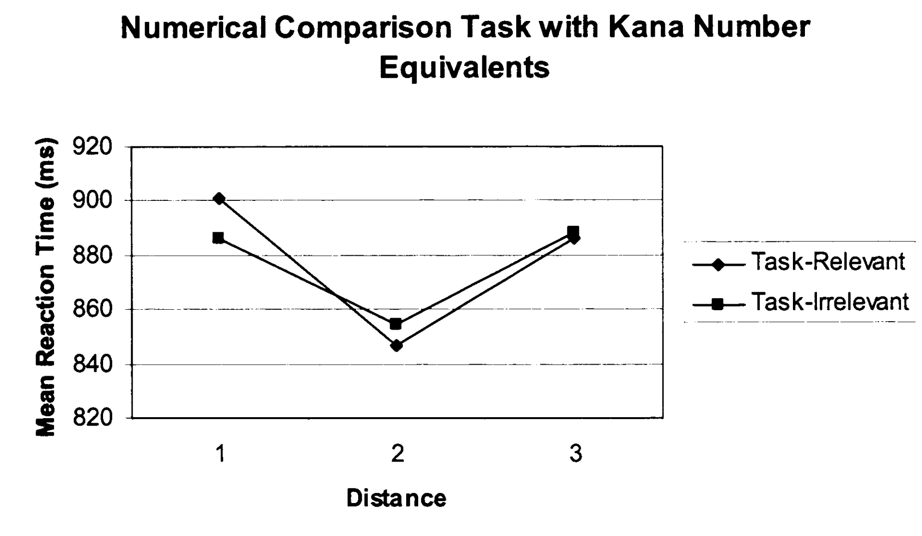


Figure 2.21 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during numerical comparisons with Kana number equivalents (Experiment 2)

For physical comparison with Kana number equivalents, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,47)} = 3.32$, *n.s.*), a non-significant main effect of distance ($F_{(2,94)} < 1$, *n.s.*), and a significant task-relevance x distance interaction ($F_{(2,92)} = 4.41$, $p < 0.050$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was significant at the linear level ($F_{(1,47)} = 8.45$, $p < 0.010$), but non-significant at the quadratic level ($F_{(1,47)} < 1$, *n.s.*). A one-way ANOVA was conducted at each level of task-relevance. Physical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,94)} = 5.78$, $p < 0.005$). Tests of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,47)} = 8.73$, $p \leq 0.005$), and a non-significant quadratic trend ($F_{(1,47)} < 1$, *n.s.*). The task-irrelevant dimension, numerical distance, showed a non-significant main effect ($F_{(2,94)} < 1$, *n.s.*). Figure 2.22 shows the reaction time patterns for physical comparison task with Kana number equivalents.

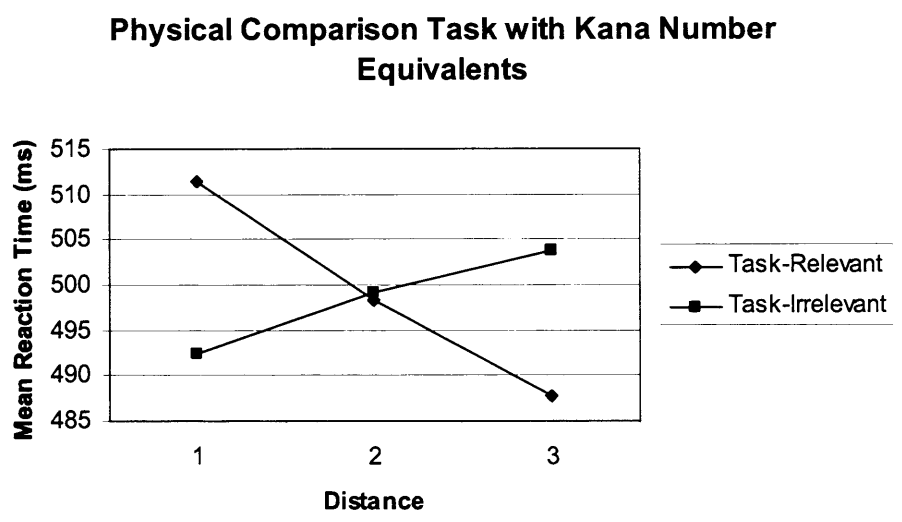


Figure 2.22 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during physical comparisons with Kana number equivalents (Experiment 2)

2.4.2.4.4 English Written Verbal Numerals

For numerical comparison with English written verbal numerals, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,46)} < 1$, *n.s.*), a significant main effect of distance ($F_{(2,80)} = 21.99$, $p < 0.001$), and a significant task-relevance x distance interaction ($F_{(2,92)} = 4.44$, $p = 0.050$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was non-significant at the linear level ($F_{(1,46)} = 2.05$, *n.s.*), but significant at the quadratic level ($F_{(1,46)} = 4.82$, $p < 0.050$). A one-way ANOVA was conducted at each level of task-relevance. Numerical distance, being the task-relevant dimension, showed a significant main effect ($F_{(2,92)} = 10.58$, $p < 0.001$). Tests of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,46)} = 7.38$, $p < 0.010$), and a significant quadratic trend ($F_{(1,46)} = 14.87$, $p < 0.001$). The task-irrelevant dimension, physical distance, showed a non-significant main effect ($F_{(2,92)} = 1.19$, *n.s.*). Figure 2.23 shows the reaction time patterns for numerical comparison task with English numerals.

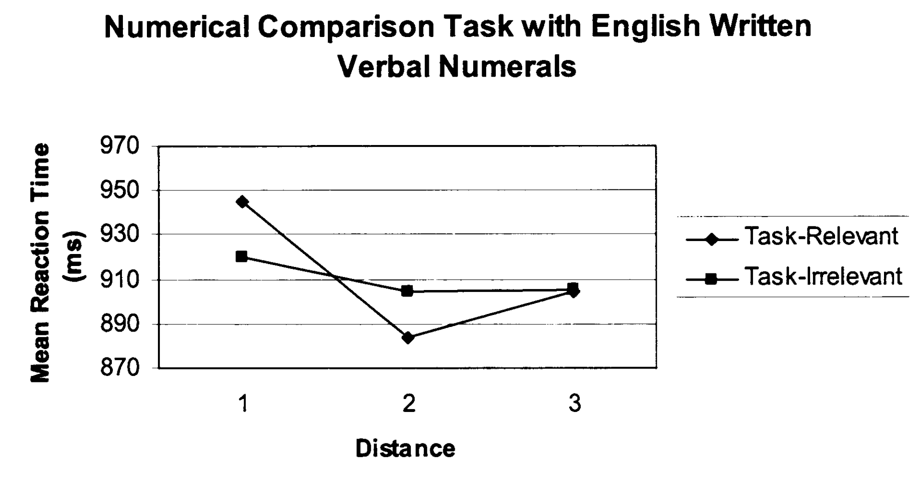


Figure 2.23 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during numerical comparisons with English written verbal numerals (Experiment 2)

For physical comparison with English written verbal numerals, the ANOVA revealed a non-significant main effect of task-relevance ($F_{(1,47)} = 2.22$, *n.s.*), a significant main effect of distance ($F_{(2,94)} = 3.28$, $p < 0.050$), and a significant task-relevance x distance interaction ($F_{(2,92)} = 4.53$, $p < 0.050$). Test of within-subjects contrasts revealed that the task-relevance x distance interaction was non-significant at the linear level ($F_{(1,47)} < 1$, *n.s.*) but significant at the quadratic level ($F_{(1,47)} = 10.42$, $p < 0.005$). A one-way ANOVA was conducted at each level of task-relevance. Physical distance, being the task-relevant dimension, showed a non-significant main effect ($F_{(2,94)} = 1.43$, *n.s.*). The task-irrelevant dimension, numerical distance, showed a significant main effect ($F_{(2,94)} = 6.39$, $p < 0.005$). Tests of within-subjects contrasts revealed a non-significant linear trend ($F_{(1,47)} < 1$, *n.s.*), but a significant quadratic trend ($F_{(1,47)} = 13.24$, $p \leq 0.001$). Figure 2.24 shows the reaction time patterns for physical comparison task with English numerals.

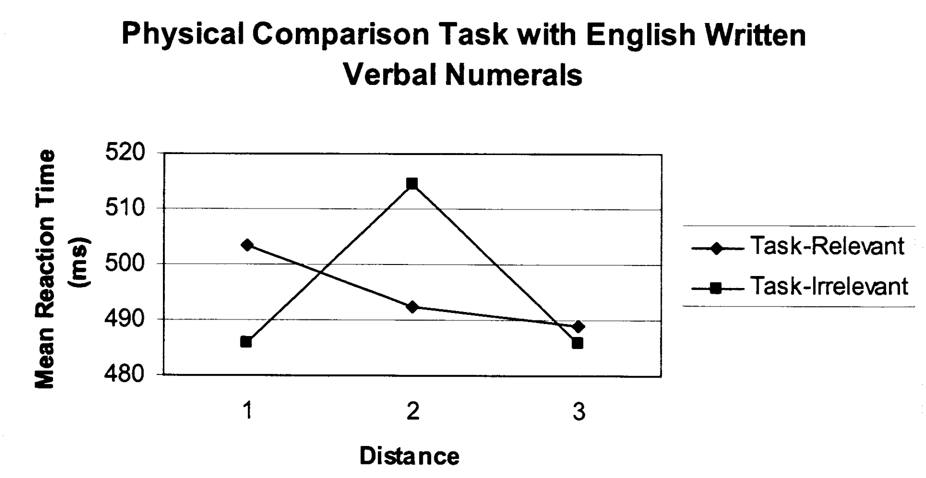


Figure 2.24 Mean reaction times (ms) at different levels of distance under task-relevant and -irrelevant conditions during physical comparisons with English written verbal numerals (Experiment 2)

2.4.2.4.5 Summary

Table 2.7 Summary of manifestation of distance effects across different stimulus types (Experiment 2)

Comparison Task	Stimulus Type	Distance Effect in the Task-Relevant Dimension (Negative Linear Trend)	Distance Effect in the Task-Irrelevant Dimension (Positive Linear Trend)
Numerical	Digit	Yes	Yes
	Kanji	Yes	Yes
	Kana	Yes†	No
	English	Yes†	No
Physical	Digit	Yes	Yes
	Kanji	Yes	No
	Kana	Yes	No
	English	No	No

†Despite the significant negative linear trend in the task-relevant dimension, the reaction time pattern was better accounted for by the significant quadratic trend. Hence, the numerical distance effect was unwarranted.

2.4.3 Discussion

Experiment 2 was designed to investigate the processing of numerical information in the two Japanese scripts: Kanji and Kana. Japanese numerals are commonly written in Kanji, an ideographic script, but rarely in Kana, a syllabic script. The present experiment, using the comparison Stroop paradigm, examined the similarities and differences in processing numerical magnitudes written in these scripts, compared to Arabic digits and numerals written in English – a language which the subjects only learned later in life. Moreover, a very first attempt was made to develop a parametric design for the comparison Stroop paradigm. In addition, two programs – one written in Cogent and the other in JavaScript – were tested for their equivalence.

2.4.3.1 Program Equivalence

Although the mean reaction time of JavaScript program users was significantly faster than that of Cogent program users, the two groups did not differ significantly on mean error rates and the factor program did not interact significantly with any other factor in both error and reaction time analyses, suggesting that patterns of mean error rates and mean reaction times were similar in the two groups.

The significant difference in mean reaction times between Cogent and JavaScript users could be attributed to individual differences – subjects who were tested using the Cogent program had not only a longer mean reaction time of 690 ms but also a larger standard deviation of 105 ms, compared to JavaScript program users' mean reaction time of 571 ms and standard deviation of 84 ms. There was no reason to suggest any inherent delay in time recording with Cogent, giving rise to a systematic error.

2.4.3.2 Comparison Tasks and Stimulus Types

Before discussing the Stroop effect, it is important to examine the differences between the two comparison tasks across the four stimulus types. There were significant differences in mean error rates and mean reaction times between the comparison tasks –7.60% vs. 1.36% and 711 ms vs. 491 ms for numerical and physical comparison tasks respectively. The differences between physical and numerical comparison tasks suggest that these two types of comparison are very different in nature – the former can be solved perceptually, whereas the latter needs to be solved conceptually. Consequently, it is unsurprising that numerical comparisons took longer to perform and were more susceptible to errors.

Mean error rates and mean reaction times showed similar patterns in each of the two tasks. For physical comparisons, the four stimulus types did not differ significantly with respect to either error rates or reaction times. However, for numerical comparisons, both measures differed significantly across stimulus types. English numerals showed both highest mean error rate and slowest mean reaction time, followed by Kana number equivalents, which were followed by Kanji numerals, and Arabic digits had the lowest mean error rate and slowest mean reaction time (note that all of these differences were statistically significant). These findings suggest that numerical comparisons were most difficult with English numerals, followed by Kana number equivalents, which were followed by Kanji numerals, and the easiest numerical comparisons occurred with Arabic digits. The significant main effects of task with respect to both mean error rates and mean reaction times suggest that physical comparisons were easier than numerical comparisons.

The pattern of error rate and reaction time may be explained in terms of subjects' familiarity with the different stimulus types. Arabic digits, used in everyday life, are familiar, and hence numerical comparisons with this stimulus type occurred relatively smoothly. Japanese numerals are commonly written in Kanji, but rarely in Kana, meaning that there was a familiarity discrepancy between these stimulus types. It is therefore unsurprising that subjects committed more errors and were slower when comparing Kana number equivalents numerically. In contrast to Kanji, English was a relatively unfamiliar language which subjects did not start to learn until the age of 13. This lack of experience was likely to lead to inaccurate and slow performance in numerical comparisons with English written verbal numerals.

2.4.3.3 The Stroop Effect

The Stroop effect, manifested as facilitation and/ or interference, was observed with all stimulus types except English numerals (see Table 2.6 for a summary).

With Arabic digits, the Stroop effect was observed as both interference and facilitation during numerical comparisons, but not during physical comparisons. However, in previous studies, the Stroop effect, manifested as interference and facilitation, was observed in both types of comparison tasks (Girelli et al., 2000;

Henik & Tzelgov, 1982; Tzelgov et al., 1992). The absence of Stroop effect in the physical comparison task could be attributed to the lack of room for improvement in terms of reaction time in non-conflict (congruent and neutral trials; given that physical comparisons could be solved perceptually in a relative short period of time even when informational conflict was present (mean reaction time was under 500 ms), there was little room for speeding up in cases where no conflict was present.

There was no evidence for a Stroop effect with English written verbal numerals. One explanation is the use of a fixed-pairs design which might have diluted any potential effect (Foltz et al., 1984). In the present experiment, only 6 numbers and 6 physical sizes were used to construct 3 levels of numerical and physical sizes. The use of a limited number of stimulus pairs might have allowed subjects to rely on a memory strategy in the comparison tasks, where subjects responded upon memory of previous trials rather than upon comparison per se (see Section 1.5.1 for details).

An alternative explanation for the absence of a Stroop effect with English numerals rests on the lack of experience with this stimulus type. English was the second language for all the subjects; although all subjects had learnt English for at least 6 years at high school, they had not achieved fluency in this language. This might have given rise to an inefficient access to numerical magnitudes. This explanation is consistent with Girelli et al.'s (2000) finding that the Stroop effect during physical comparisons (where numerical magnitude was the task-irrelevant dimension) did not emerge until children had a firm understanding of numerical magnitude, at around age 8 for Italian children.

Note that the above explanations are not incompatible with one another. On the other hand, it is perhaps the interaction of these factors which has led to an absence of the Stroop effect with English written verbal numerals; it might be possible to observe a weak version of the effect if a repeated-set design were employed; however, given that English was the second and a relatively unfamiliar language for the subjects, the use of a fixed-pairs design might have resulted in an absence of such an effect.

The inefficient access to semantics with English numerals may also be viewed as a weak pathway in terms of Cohen et al.'s (1990) idea of the strength of a process. According to these authors, the stronger the pathway of a process is, the less attention it requires, and the more likely that it produces interference. English as a second language learnt late in life could mean a weak pathway to semantics, and hence less potency in interfering with the task-relevant dimension, physical size.

With respect to the two Japanese scripts, only facilitation was observed with Kanji numerals in both numerical and physical tasks, but with Kana number equivalents, interference was observed in both tasks and facilitation only in the numerical task. In a thorough review on the Stroop phenomenon, MacLeod (1991) pointed out that interference was virtually always substantially larger than facilitation (MacLeod, 1991). Thus, the finding that facilitation was observed without interference with Kanji numerals in both tasks was intriguing. The patterns of the Stroop effect with regard to the two Japanese scripts are further discussed in Section 2.4.3.4.

2.4.3.4 Degree of Information Processing

The use of a parametric design has allowed the investigation of the degree of interference exerted by stimuli from the task-irrelevant dimension of different levels of numerical and physical distance; the analyses were conducted on mean reaction times of incongruent trials only. In any given comparison task, a significant negative linear trend of the task-relevant distance (i.e., the classic numerical distance effect) would indicate that information processing was refined in the task-relevant dimension, whereas a significant positive linear trend of the task-irrelevant distance (i.e., the reversed numerical distance effect) would indicate refined information processing in the task-irrelevant dimension.

During numerical comparisons with Arabic digits, a distance effect (indicated by a significant negative linear trend) was observed in the task-relevant dimension –

numerical magnitude. In other words, the classic numerical distance effect was observed, replicating previous research findings (e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977). In addition, a reversed distance effect was observed in the task-irrelevant dimension, physical size. The latter had never been previously reported.

The equivalent effects were observed during physical comparisons with Arabic digits, i.e., a physical distance effect when physical size in the task-relevant dimension (the classic distance effect) and a reversed numerical distance effect when numerical magnitude was the task-irrelevant, the latter replicating findings of Henik and Tzelgov's (1982) and Girelli et al. (2000).

It is important to point out that there was an absence of a significant interference effect during physical comparisons with Arabic digits. Despite this, the reversed numerical distance effect provides strong evidence for autonomous processing of numerical magnitude under task-irrelevant conditions. These findings suggest that the Stroop effect may not be a sensitive enough measure for autonomous processing. Instead, distance effects provide a more precise measure for the degree of information processing in both task-relevant and -irrelevant conditions, and in the latter provide a more sensitive measure for autonomous processing. In summary, numerical magnitudes and physical sizes of Arabic digits were processed in a refined fashion both when they were task-relevant and -irrelevant.

The Stroop effect manifested only as facilitation with Kanji numerals in both numerical and physical tasks. Further analyses revealed a significant physical distance effect in the physical task, a significant numerical distance effect in the numerical task and, and more importantly, a reversed physical distance effect in the numerical task. The absence of an interference effect and the significant reversed physical distance effect again suggest that the Stroop effect as a measure of autonomous processing was insensitive. During numerical comparisons, not only the numerical magnitudes, but also the task-irrelevant physical sizes, were processed in a refined fashion. On the other hand, the absence of an interference effect and the absence of a reversed numerical distance effect during physical comparisons confirm that numerical magnitudes of Kanji numerals did not interfere with the task-relevant physical comparison task.

Interference was observed with Kana number equivalents in both numerical and physical tasks. However, the absence of a reversed distance effect in the two corresponding task-irrelevant dimensions suggests that although information from the task-irrelevant dimension interfered with the relevant task, the processing of such information was coarse. Even when task-relevant, Kana number equivalents did not exhibit a significant numerical distance effect (although there was a significant negative linear trend, the reaction time pattern was better accounted for by a quadratic trend). This could have resulted from the lack of familiarity of the stimuli. Numerical comparisons of these stimuli were relatively difficult compared to Arabic digits and Kanji numerals. This could mean that congruent and neutral trials had little advantage over incongruent trials, and hence the apparent lack of refined processing. An alternative explanation is that since these stimuli were unfamiliar and a fixed-pairs design was used, subjects might have adopted a memorial strategy to cope with the relatively difficult numerical comparisons and made judgements upon their responses to previous trials (see Section 1.5.1). This would have diluted any potential Stroop effect and numerical distance effect. In summary, only physical sizes of the Kana stimuli under task-relevant condition was processed in a refined fashion. Numerical magnitudes of these stimuli were not processed in a refined fashion regardless of task-relevance.

The total absence of a significant distance effect has confirmed that the English written verbal numerals were not processed in refined fashion by the Japanese subjects under both task-relevant and -irrelevant conditions. (Note that during numerical comparisons, even though reaction time pattern showed a significant negative linear trend with numerical distance, it was better accounted for by a quadratic trend.)

2.4.3.5 Measuring Autonomous Information Processing

The current results have spelt out the limitations in using the Stroop effect to demonstrate autonomous processing. Interference – one of the component effects of the Stroop phenomenon – has often been used to indicate autonomous processing of task-irrelevant information, but current findings have shown that it is an insensitive measure for this purpose. Only when interference is coupled with a significant reversed distance effect in the task-irrelevant dimension can one confirm that the task-irrelevant information is processed autonomously in a refined fashion.

Moreover, the reversed numerical distance effect observed with Arabic digits in the current experiment was determined by a linear trend established from specifying polynomial contrasts in the analysis, instead of being implicated from a mere difference between numerically close and numerically distant conditions (Girelli et al., 2000; Henik & Tzelgov, 1982). The statistical analysis used here has allowed the pinpointing of the change in mean reaction time pattern with numerical distance, providing powerful evidence for refined autonomous processing of numerical magnitude under task-irrelevant condition.

2.4.3.6 Effects of Writing System

The present experiment was designed to investigate the processing of numerical information in the two Japanese scripts, Kanji and Kana. Japanese numerals are commonly written in Kanji (an ideographic script), but rarely in Kana (a syllabic script). Similarities and differences in processing numerical magnitudes written in these scripts were examined and compared to Arabic digits and numerals written in English – a language which the subjects only learned later in life.

Even though Kana number equivalents are not parts of everyday usage, the processing of such stimuli has revealed certain similarities to that of Kanji numerals. Both stimulus types exhibited the Stroop effect, manifested as facilitation and/ or interference. However, when task-relevant, Kanji numerals were processed in a refined manner, but not Kana number equivalents. In contrast, these stimulus types behaved similarly under task-irrelevant conditions; unlike Arabic digits, there was no evidence to support the idea that numerical magnitudes of these stimuli were processed in a refined fashion when they were task-irrelevant.

Differences in terms of the Stroop effect and distance effects between the two Japanese scripts result primarily from the discrepancy in familiarity of these stimuli. Japanese numerals are commonly written in Kanji, but rarely in Kana. Thus, comparisons with the latter were novel and in the case of numerical comparisons, more difficult. In terms of Cohen et al.'s (1990) idea of the strength of a process, Kanji numerals would have a stronger pathway to semantics due to its familiarity in everyday use, whereas Kana number equivalents would have a weaker pathway to semantics due to its rare appearance in everyday life. This account is consistent and satisfactory with the present findings regarding the presence and absence of a classic numerical distance effect with Kanji and Kana stimuli respectively during numerical comparisons.

The absence of a Stroop effect with English numerals may be explained in terms of a fixed-pairs design (Foltz et al., 1984; see also Section 1.5.1) and in terms of subjects' lack of fluency in this language. The latter relates to Cohen et al.'s (1990) idea of the strength of a process. In the case of English being the second language of the subjects, the pathway to semantics would be weak, and hence the stimuli would be less capable in producing a Stroop effect.

The different patterns in terms of the Stroop effect and distance effects across writing systems should not be attributed to the mere differences in orthography. According to Yin and Butterworth (1992), different types of writing scripts have the same broad underlying cognitive architecture (see Section 2.3.3.1). Other factors such as familiarity with the stimuli, fluency in the language of the stimuli, and the strength of pathway to semantics play important roles in determining the degree of information processing of numerical magnitudes.

In summary, autonomous processing of numerical magnitudes was only evident with Arabic digits.

3 Investigation into the Autonomous Processing of Numerical Magnitudes with a Fully Parametric Design

3.1 *Experiment 3a: Behavioural Experiment*

3.1.1 Introduction

The comparison Stroop paradigm has been widely used to investigate numerical information processing (e.g., Besner & Coltheart, 1979; Foltz et al., 1984; Girelli et al., 2000; Henik & Tzelgov, 1982; Yurko & Hinrichs, 1978). In a typical experiment, subjects would be asked to judge which of the number pair is larger (or smaller) either in numerical magnitude or in physical size (e.g., Girelli et al., 2000; Henik & Tzelgov, 1982).

There is some inconsistency as to whether the Stroop effect occurs in both numerical and physical comparisons. Henik and Tzelgov (1982) observed the Stroop effect only during numerical comparisons, but not physical comparisons. However, Girelli et al. (2000) reported the effect (manifested as interference) in both numerical and physical comparison tasks. The interference observed during physical comparisons has been used to implicate autonomous processing of numerical magnitudes. Results in Experiments 1 and 2 have, however, suggested that interference as a measure of autonomous processing may not be sensitive enough. Instead, results have shown that autonomous processing of numerical magnitude is better reflected by the reversed numerical distance effect under task-irrelevant condition, i.e., during physical comparison task. The effect was first observed by Henik & Tzelgov (1982) and replicated by Girelli et al. (2000), but Rubinsten et al. (2002) failed to replicate it. The reversed numerical distance effect never received much attention from researchers, but the current thesis argues that it can be used an important indicator of semantic processing under task-irrelevant conditions.

The inconsistent findings with regard to the reversed numerical distance effect may be explained in terms of inappropriate experimental design. In order to investigate distance effects, it is important to vary systematically the levels of distance in the two dimensions, namely numerical magnitude and physical size. In fact, very few attempts have been made to vary both of these dimensions, and they were not methodologically successful. For example, Rubinsten et al. (2002) used three levels of numerical distance and two levels of physical distance (where the physical sizes of the digits differed either by 1 mm or 2 mm) in their number Stroop experiment and observed a numerical distance effect when numerical magnitude was the task-relevant dimension but not when it was task-irrelevant. The authors reported a physical distance effect but did not test for any physical distance effect when physical size was the task-irrelevant dimension. It was also unclear why different numbers of levels were employed with respect to numerical distance and physical distance.

More recently, Pinel et al. (2004) varied both numerical distance and physical distance, but the stimuli were grouped in the analyses, so that the factor distance consisted of only two levels (distant and close pairs). With such a design, Pinel et al. (2004) failed to observe any distance effect in either numerical magnitude or physical size when they were task-irrelevant.

In Experiment 2, the first attempt was made to balance the two dimensions being investigated. Three levels of distance were used in both the numerical and the physical dimensions. More importantly, instead of merely testing the difference between numerically distant and numerically close trials, polynomial contrasts were specified, allowing linear trends to be statistically established (see Experiment 2). The reversed numerical distance effect revealed by the latter method provides statistically more powerful evidence for the autonomous processing of numerical magnitudes under task-irrelevant conditions.

The current experiment aimed to replicate the findings of Experiment 2 using the full range of single digits from 1 to 9 (instead of only 6 in Experiment 2) and 9 physical sizes to construct 4 levels of distance in each of the two dimensions.

Two pilot studies were carried out to match the dimensions in terms of difficulty prior to the current experiment.

A physical distance effect in the physical comparison task was predicted based on the finding that distance effect is a robust phenomenon in perceptual comparisons (e.g., Cattell, 1902). When physical size is the task-irrelevant dimension (i.e., in a numerical comparison task), it would be intuitive to expect a reversed physical distance effect. If the digits being compared are highly discriminable (i.e., very different in their physical sizes), then it would be faster to make a physical size comparison than if they are of similar sizes – this is the classic distance effect. However, physically distant pairs, with their salient perceptual differences, would exert a stronger Stroop effect on numerical comparisons than physically close pairs; the closer the physical sizes, the closer a congruent or incongruent pair appears to be a neutral pair. For example, selecting the numerically larger item will be slower for 3 5 than 3 5 since it will be quicker to resolve the physical size difference in the first case.

The above predictions fit those made by Algom et al. (1996) based on Melara and Mounts' (1993) notion of discriminability, which specifies the psychological difference separating two stimulus values along a dimension. In terms of the comparison Stroop paradigm, discriminability is matched if the values along the numerical dimension (i.e., magnitudes) are as different psychologically as the values along the physical dimension (i.e., physical sizes) (Algom et al., 1996). As Melara and Mounts (1993) showed, Stroop interference was malleable, with the more discriminable dimension causing a failure of selective attention to the less discriminable dimension, but not vice versa. Thus, physically distant incongruent pairs would disrupt numerical comparison performance more than physically close pairs, and similarly, numerically distant incongruent pairs would interfere more with physical size comparisons than numerically close pairs. Indeed, in Pansky and Algom's (1999) number Stroop experiment, they found that when the task-irrelevant dimension was more discriminable, a sizeable Stroop effect affected performance on the task-relevant dimension, but when it was less discriminable, the Stroop effect was considerably weaker. These findings were consistent with the reversed numerical distance effect in the task-irrelevant channel (i.e., the physical comparison task) observed in Henik and Tzelgov's (1982) and Girelli et al.'s (2000) experiments, and in Experiment 2 of the current thesis, on the assumption that distant pairs are more discriminable than close pairs, thus exerting a stronger size congruity effect on the task-relevant dimension.

In summary, by employing a parametric design of the comparison Stroop paradigm, the present experiment aimed to replicate (1) the basic finding of a Stroop effect in both numerical magnitude and physical size comparison tasks (e.g., Girelli et al., 2000) and (2) the autonomous processing of numerical magnitudes. The latter would be implicated by a reversed distance effect in the task-irrelevant condition, i.e., in the physical comparison task.

3.1.2 Methods

3.1.2.1 Tasks

The two tasks were numerical magnitude comparison and physical size comparison. Subjects had to select, via a key response (by pressing the left "F" key with their left index finger or the right "J" key with their right index on a qwerty keyboard) the larger number of the pair in numerical magnitude or in physical size accordingly. Their reaction times and responses were recorded.

The tasks were computer-based. A program written in Cogent (running on a MATLAB Version 6.1 platform) was used. It obtained inputs from and recorded outputs (reaction times and key responses) onto text files.

3.1.2.2 Stimuli

In each trial, two digits appeared in white on a black background in the middle of a 15" TFT screen. Each presentation lasted 1000 ms, and there was an interval of 2000 ms before the subsequent trial. The digits were approximately 3.50 cm apart. Subjects sat approximately 45 cm away from the screen.

The stimuli were Arabic digits (1 to 9) in font Arial. Each digit might appear in one of the nine different sizes – the smallest was 1.00 cm and the largest 2.60 cm in height, with increments of 0.20 cm in between; the visual angles were approximately 1.3°, 1.5°, 1.8°, 2.0°, 2.3°, 2.5°, 2.8°, 3.1°, and 3.6°, with an average ratio of 1.1 between adjacent sizes. These particular physical sizes were selected following two pilot studies which aimed to match the two dimensions, namely numerical magnitudes and physical sizes, by difficulty level/ error rate (see Section 33.1.3).

There were 3 experimental conditions: congruent, when the numerically larger digit was physically larger (e.g., 2 6); incongruent, when the numerically larger digit was physically smaller (e.g., 2 6); and neutral, when the two digits were of the same physical size in the numerical comparison task (e.g., 2 6) or when the same digit appeared in different sizes in the physical comparison task (e.g., 2 2).

Congruent and incongruent trials were varied along two dimensions, namely task-relevant distance and task-irrelevant distance. In the numerical comparison task, the task-relevant distance was the numerical distance between the two digits presented and the task-irrelevant distance was the physical distance between them, whereas in the physical comparison task, physical distance was the task-relevant dimension and numerical distance was the task-irrelevant dimension. Neutral trials, on the other hand, varied only along one dimension in each task – in the numerical task, trials varied only in numerical distance, whereas in the physical task, trials varied only in physical distance.

Four numerical distances (1, 2, 3, and 4) and four physical distances (0.20 cm, 0.40 cm, 0.60 cm, and 0.80 cm in height) were used. The pairs used were: 2 3, 3 4, 6 7, 7 8 (for numerical distance of 1); 1 3, 2 4, 6 8, 7 9 (for numerical distance of 2); 1 4, 2 5, 5 8, 6 9 (for numerical distance of 3); 1 5, 2 6, 4 8, 5 9 (for numerical distance of 4). Each of these pairs was presented twice in physical sizes belonging to each of the 4 physical distance groups, once as congruent and once as incongruent trial, making a total of 64 trials of each type. The equivalent pairs were used in the physical comparison task, e.g., for physical size of 1 unit (0.20 cm), the pair would consist of a number in size 2 (1.20 cm) and another in size 3 (1.40 cm). In the numerical comparison task, the neutral trials consisted of the same 16 pairs listed above, each appearing 4 times in physical sizes of (1.20 cm, 1.60 cm, 2.00 cm, and 2.40 cm). In the physical comparison task, the pairs used were: 1 1, 2 2, 3 3, 4 4, 6 6, 7 7, 8 8, 9 9; each pair appeared twice in physical sizes belonging to each of the 4 physical distance groups, once as congruent and once as incongruent trial, making a total of 64 neutral trials.

The total number of trials in each task was 192, half of which had the correct response appearing on the left, and the rest on the right. Stimuli were presented in a pseudorandom order with the following constraints to avoid carryover effects: (1) the same digit or the same physical size did not appear in consecutive trials, (2) the correct response did not appear on the same side (left or right) for more than three consecutive trials, (3) the experimental condition (i.e., congruent, incongruent, and neutral) was not the same for more than two consecutive trials, and (4) the same distance (numerical or physical) was not the same for more than three consecutive trials.

3.1.2.3 Subjects

There were 20 right-handed subjects (8 males and 12 females), age ranged 18 to 35 (mean = 23.7 years, standard deviation = 4.2 years). They performed two tasks (numerical and physical comparison tasks); half of them participated in the numerical task first, and the other half the physical task first. All subjects had normal or corrected-to-normal eyesight.

3.1.3 Results

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 2000 ms after the stimulus offset. Subjects who made more than 5% errors were removed from further analyses. Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were also removed.

ANOVAs were used to analyse the mean error rates and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

A 2 x 3 repeated-measures ANOVA was conducted on mean error rates. The factors were task (numerical comparison task and physical comparison task) and congruity (congruent, neutral, and incongruent). The ANOVA revealed a non-significant main effect of task ($F_{(1,15)} = 3.63$, *n.s.*), a non-significant main effect of congruity ($F_{(1,22)} = 1.59$, *n.s.*), and a non-significant task x congruity interaction ($F_{(2,30)} < 1$, *n.s.*). During numerical comparisons, the mean error rates were 4.01%, 4.10%, and 5.47% for congruent, neutral, and incongruent trials respectively (see Figure 3.1). During physical comparisons, the mean error rates were 5.86%, 5.27%, and 6.25% for congruent, neutral, and incongruent trials respectively (see Figure 3.1).

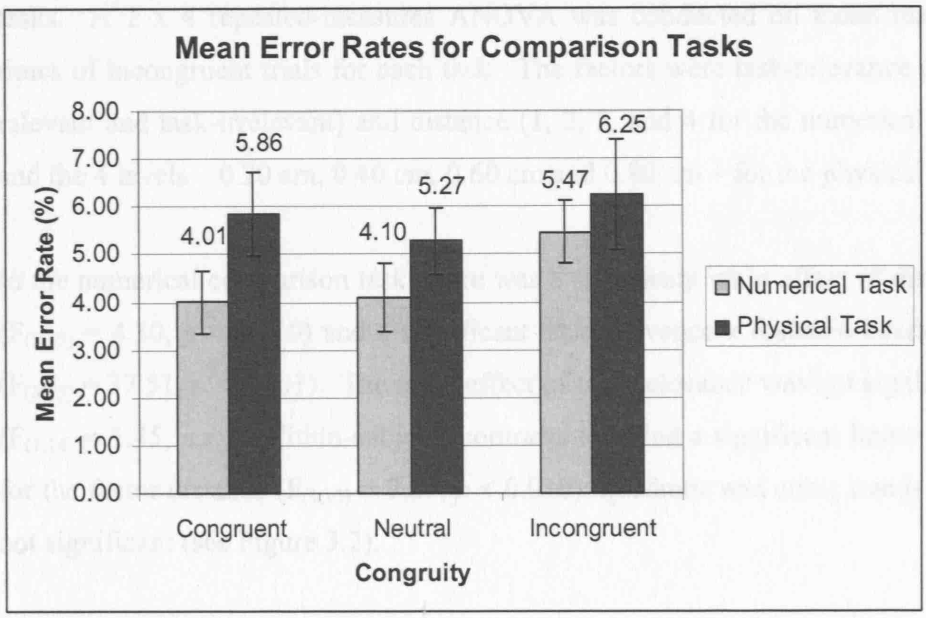


Figure 3.1 Mean error rates for numerical and physical comparison tasks (Experient 3a)

A 2 x 3 x 2 mixed design ANOVA was conducted on mean reaction times. The factors were: task (numerical comparison task and physical comparison task), congruity (congruent, neutral, and incongruent), and order (numerical comparison task first, then physical comparison task, and physical comparison task first, then numerical comparison task). Order was the only between-subjects factor.

The ANOVA revealed a significant main effect of task ($F_{(1,14)} = 15.65, p \leq 0.001$) – physical comparisons were significantly faster than numerical comparisons (513 ms and 572 ms respectively), a significant main effect of congruity ($F_{(2,28)} = 125.67, p < 0.001$), but the main effect of order was not significant ($F_{(1,14)} < 1$, n.s.). Moreover, there was no significant interaction.

Tests of within-subjects contrasts revealed significant differences in mean reaction times between congruent and neutral trials ($F_{(1,14)} = 31.64, p < 0.001$), and between neutral and incongruent trials ($F_{(1,14)} = 105.02, p < 0.001$). The mean reaction times were 519 ms, 533 ms, and 575 ms respectively.

Distance effects were tested separately for numerical and physical comparison tasks. A 2 x 4 repeated-measures ANOVA was conducted on mean reaction times of incongruent trials for each task. The factors were task-relevance (task-relevant and task-irrelevant) and distance (1, 2, 3, and 4 for the numerical task, and the 4 levels – 0.20 cm, 0.40 cm, 0.60 cm and 0.80 cm – for the physical task).

In the numerical comparison task, there was a significant main effect of distance ($F_{(3,45)} = 4.30, p < 0.010$) and a significant task-relevance x distance interaction ($F_{(3,45)} = 37.51, p < 0.001$). The main effect of task-relevance was not significant ($F_{(1,14)} = 1.45$, *n.s.*). Within-subjects contrasts revealed a significant linear trend for the factor distance ($F_{(1,15)} = 9.65, p < 0.010$); quadratic and cubic trends were not significant (see Figure 3.2).

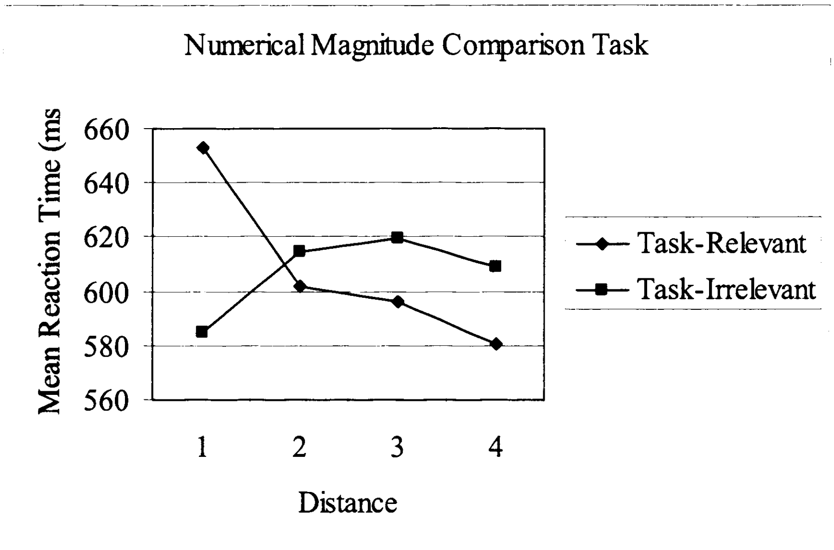


Figure 3.2 Mean reaction times (ms) for incongruent trials during numerical comparison task (Experiment 3a)

When numerical magnitudes were task-relevant, the main effect of distance was significant ($F_{(3,45)} = 28.97$, $p < 0.001$) and the factor distance showed a significant negative linear trend ($F_{(1,15)} = 86.03, p < 0.001$) (see Figure 3.2).

When physical sizes were task-irrelevant, the main effect of distance was significant ($F_{(3,45)} = 5.99, p \leq 0.002$) and the factor distance showed a significant positive linear trend ($F_{(1,15)} = 4.81, p < 0.050$) (see Figure 3.2).

In the physical comparison task, there was a significant main effect of distance ($F_{(2,24)} = 12.97, p < 0.001$) and a significant task-relevance x distance interaction ($F_{(1,22)} = 21.47, p < 0.001$). The main effect of task-relevance was not significant ($F_{(1,15)} = 1.82$, *n.s.*). Within-subjects contrasts revealed a significant linear trend for the factor distance ($F_{(1,15)} = 22.31$, $p < 0.001$); quadratic and cubic trends were not significant (see Figure 3.3).

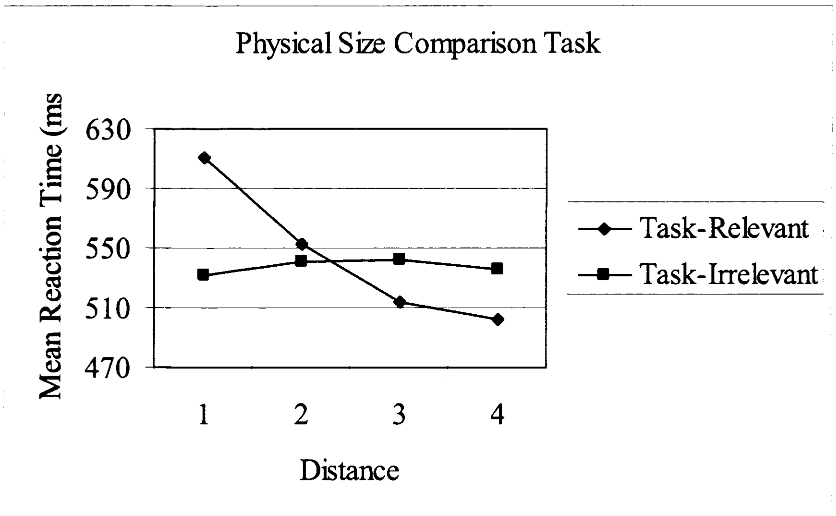


Figure 3.3 Mean reaction times (ms) for incongruent trials during physical comparison task (Experiment 3a)

When physical sizes were task-relevant the main effect of distance was significant ($F_{(2,28)}$ = 83.16, p < 0.001) and the factor distance showed a significant negative linear trend ($F_{(1,15)}$ = 155.88, p < 0.001) (see Figure 3.3).

When numerical magnitudes were task-irrelevant, the main effect of distance was not significant ($F_{(1,20)}$ = < 1, *n.s.*) and the linear trend for factor distance was not significant ($F_{(1,15)}$ < 1, *n.s.*) (see Figure 3.3).

3.1.4 Discussion

The present study aimed to examine the Stroop effect when processing numerical information by employing two comparison tasks – namely numerical magnitude and physical size comparison tasks.

Mean error rates for the two comparison tasks did not differ significantly, suggesting that the two tasks were matched closely in terms of difficulty level. However, the Stroop effect did not emerge in either task when considering mean error rates. Hence, error rates could not be considered a sensitive indicator for the Stroop effect. Consequently, the rest of the discussion will focus on mean reaction time analyses.

Mean reaction time for physical comparisons was significantly faster than that for numerical comparisons. The Stroop effect, manifested as interference and facilitation, was observed in both tasks (replicating previous findings, e.g., Girelli et al., 2000; Henik & Tzelgov, 1982), with the former being more prominent than the latter (consistent with previous findings, see review, MacLeod, 1991).

The use of a parametric design has allowed the investigation of the degree of interference exerted by stimuli of different numerical and physical distances. The analyses were conducted on mean reaction times of incongruent trials only, where there were informational conflicts. As predicted, a distance effect (revealed by a significant negative linear trend) was always observed in task-relevant dimensions, i.e., numerical distance effect in numerical comparison task and physical distance effect in physical comparison task. These findings indicate that information processing was refined under task-relevant conditions.

When task-irrelevant, a reversed physical distance effect (revealed by a significant positive linear trend) was observed, but no reversed numerical distance effect was observed. Note that the positive linear trend of the physical distance effect in numerical comparisons was not as steep as the negative linear trend of physical distance in physical comparisons. Such findings suggest that autonomous information processing in the task-irrelevant dimension was, although refined, relatively less detailed than that in the task-relevant dimension.

The distance effects observed in task-relevant dimensions are consistent with previous findings: shorter reaction time was required in numerical comparisons if the two numbers had a larger numerical distance than if they had a smaller numerical distance (e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977; see also Experiment 2 of the current thesis) and in physical comparisons if they had a larger physical distance than if they had a smaller physical distance (replicating finding of Experiment 2). Furthermore, a reversed physical distance effect was observed when physical sizes were the task-irrelevant dimension (in numerical comparison task), also replicating the findings of Experiment 2.

The absence of a reversed numerical distance effect when numerical magnitudes were task-irrelevant is inconsistent with previous findings obtained by Henik and Tzelgov (1982), Girelli et al. (2000), and that of Experiment 2 of the present thesis. One possible explanation for the inconsistency is that the physical comparison task in the current experiment was more difficult than that used by previous researchers and that in Experiment 2. Since nine sizes were used to construct four levels of physical distance in the current experiment and that adjacent sizes differed by only 0.20 cm (or between 0.2° and 0.5°), physical comparisons were relatively difficult. This means that subjects were likely to concentrate more on the task-relevant physical dimension and this might have helped them to ignore more successfully the task-irrelevant numerical magnitudes, and hence the absence of a reversed numerical distance effect.

The absence of a reversed numerical distance effect in the physical comparison task might have resulted from low discriminability of the physical sizes. Consequently, modifications were made in Experiment 3b to disambiguate them by increasing the absolute physical sizes of the stimuli.

3.2 Experiment 3b: Functional Magnetic Resonance Imaging (fMRI) Experiment

3.2.1 Introduction

The experimental design of Experiment 3b was exactly the same as that of Experiment 3a, except that the physical sizes were increased to ensure that the stimuli were physically discriminable.

In addition to replicating the basic finding of a Stroop effect in both numerical magnitude and physical size comparison tasks (e.g., Girelli et al., 2000), the present experiment was designed to investigate the autonomous processing of numerical magnitudes. It aimed also to replicate the classical distance effect in task-relevant conditions: a numerical distance effect in the numerical task (e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977; see also Experiment 2 of the present thesis) and a physical distance effect in the physical task (see Experiment 2), and to replicate the reversed distance effect under task-irrelevant conditions – a reversed numerical distance effect in the physical task (e.g., Girelli et al., 2000; Henik & Tzelgov, 1982; see also Experiment 2) and a reversed physical distance effect in the numerical task (see Experiment 2).

More importantly, using the functional Magnetic Resonance Imaging (fMRI) technique, the present experiment was designed to investigate (1) the involvement of the parietal lobes in the processing of numerical information, and (2) the brain areas involved in conflict resolution and error commission.

There is already considerable evidence as to where semantic processing of numbers takes place in the brain. Pinel et al. (2001) have described two distinct stages of numerical processing, namely identification and semantic processing. Early visual cortices are implicated in the identification process: the ventral occipitotemporal areas are activated bilaterally by the visual shapes of Arabic numerals (Dehaene & Cohen, 1995). Word identification is thought to be strictly left-lateralized and to rely on the left "visual word form area", a region of the left fusiform gyrus which is involved in the invariant recognition of visual words (Cohen & Dehaene, 2004; Cohen, Dehaene, Naccache, Lehéricy, Dehaene-Lambertz, Hénaff, et al., 2000; see also Price & Devlin, 2003 who challenged the idea of specialization for visual word form representations). However, Pinel and colleagues have provided evidence that the right fusiform gyrus is implicated in the identification of Arabic numerals (Pinel, Le Clec'H, van de Moortele, Naccache, Le Bihan, & Dehaene, 1999; Pinel et al., 2001).

Several studies implicate the parietal lobes in supporting a notation-independent semantic representation of quantities or magnitudes (see Dehaene & Cohen, 1995 for review), where activity is modulated by numerical distance (e.g., Pinel et al., 2001, 2004). In particular, Pinel et al. (2004), employing a number Stroop paradigm, reported that during the numerical comparison task, the numerical distance effect was associated with enhanced bilateral activity of the horizontal segment of the intraparietal sulci and the left precentral gyrus. In contrast, during physical size comparisons with number stimuli, correlates of the physical distance effect were found predominantly in the right hemisphere, in particular the right prefrontal and occipital cortices and much of the right intraparietal sulcus. Enhanced activity in bilateral regions of anterior intraparietal sulcus was associated with both numerical magnitude and physical size distance effects.

Position emission tomography (PET) has been used to examine the cerebral networks involved in number comparison and findings converge to a fronto-parietal network. In particular, Dehaene, Tzourio, Frak, Raynaud, Cohen, Mehler, and Mazoyer (1996) reported that both comparison and multiplication activated the left and right lateral occipital cortices, the left precentral gyrus, and the supplementary motor area, and relative to multiplication, comparison yielded superior activity in the right superior temporal gyrus, the left and right middle temporal gyri, the right superior frontal gyrus, and the right inferior frontal gyrus. Similarly, Pesenti, Thioux, Seron, and De Volder (2000) reported a fronto-parietal network for comparison and simple addition involving mainly the left intraparietal sulcus, the superior parietal lobule and the precentral gyrus; comparison also activated the right superior parietal lobe.

Specific brain regions have been implicated in processing and resolution of informational conflict from task-irrelevant features. In particular, anterior cingulate cortex (ACC) activity has been reported in colour-word Stroop tasks where subjects have to name the colour of a colour word, which may be conflicting (e.g., the word GREEN printed in red) (e.g., Bench, Frith, Grasby, Friston, Paulesu, Frackowiak, & Dolan, 1993; Carter, Mintun, & Cohen, 1995; Derbyshire, Vogt, & Jones, 1998; George, Ketter, Parekh, Rosinsky, Ring, Casey, Trimble, Horwitz, Herscovitch, & Post, 1994; Pardo, Pardo, Janer, & Raichle, 1990). The ACC is also recruited during other cognitive interference tasks (Taylor, Kornblum, Minoshima, Oliver, & Koeppe, 1994), suggesting involvement of the ACC in tasks that require subjects to resolve processing conflicts between competing information streams.

However, others have suggested that the ACC is involved in motor preparation processes (e.g., Zysset et al., 2001) and in error commission (e.g., Carter, Beaver, Barch, Botvinick, Noll, Cohen, 1998; Critchley, Tang, Glaser, Butterworth, & Dolan, 2005; Kiehl, Kiddle, Hopfinger, 2000; Menon, Adleman, White, Glover, & Reiss, 2001; for a review, see also Botvinick, Braver, Barch, Carter, & Cohen, 2001).

Recent research has suggested that inferior frontal regions are responsible for conflict resolution. In particular, Zysset et al. (2001), who modified the traditional colour-word Stroop paradigm by incorporating a matching process (so that subjects made responses via key presses) to disentangle response preparation from the interference process, observed no substantial activation in the ACC when contrasting the incongruent and neutral conditions, arguing that "the ACC is not specifically involved in interference processes" but "seems rather involved in motor preparation processes" (for overview on anterior cingulate's involvement in motor execution, see Passingham, 1996; Picard & Strick, 1996; Posner & Digirolamo, 1998, and for details of Zysset et al.'s experiment, 2001, see Section 6.2.1). Instead, they observed enhanced activity in regions along the left inferior sulcus (IFS) during conflict trials compared to non-conflict (neutral and congruent trials) and concluded that "regions along the IFS appear to be involved in solving interference effect and task management."

Similar to Stroop tasks, go/no-go and stop-signal tasks require subjects to perform speeded responses on "go" trials and to inhibit their response on "no-go" or "stop" trials. Such response inhibition has been reported to activate regions along the right inferior frontal gyrus in neuroimaging studies (e.g., Bunge, Dudukovic, Thomason, Vaidya, & Gabrieli, 2002; Durston, Thomas, Yang, Uluğ, Zimmerman, & Casey, 2002; Garavan, Ross, & Stein, 1999; Konishi, Nakajima, Uchida, Kameyama, & Miyashita, 1999; Konishi, Nakajima, Uchida, Sekihara, & Miyashita, 1998; Menon et al., 2001; Rubia, Overmeyer, Taylor, Brammer, Williams, Simmons, Andrew, & Bullmore, 1999). In some studies, enhanced activity was observed bilaterally (e.g., Menon et al., 2001).

The current experiment aimed to investigate the involvement of the parietal lobes in the processing of numerical information and to examine brain regions associated with conflict resolution. In the experiment, conflict trials refer to incongruent trials where interference was present, and non-conflict trials refer to both neutral and congruent trials where no interference was present. Such a distinction was made on the basis that incongruent trials require additional attentional demands compared to both neutral and congruent trials (Milham, Erickson, Banich, Kramer, Webb, Wszalek, & Cohen, 2002; see also Carter et al., 1995; Posner & DiGirolamo, 1998, for similar explanations).

Enhanced activity was predicted in bilateral parietal regions during the numerical comparison task compared to the physical comparison task, and these regions would be examined with regard to numerical and physical distance. Frontal regions such as the left precentral gyrus and the right superior and inferior frontal gyri might also be involved during numerical comparisons.

On the basis of autonomous processing of numerical information, it was predicted that parietal activities would not be modulated by task-relevance. In other words, activation would be the same whether the dimension was relevant to the task or not. On the other hand, it was predicted that informational conflict would modulate activation in the inferior frontal regions, though not necessarily in the ACC.

3.2.2 Methods

3.2.2.1 Tasks

The two tasks were numerical magnitude comparison and physical size comparison tasks. Subjects had to select the larger number numerically or physically according to the task requirement. Subjects responded by pressing the left or right button (with their left or right thumb respectively) to indicate the side of the larger relevant attribute. Their reaction times and responses were recorded. A program written in Cogent (running on a MATLAB Version 6.1 platform) was used.

3.2.2.2 Stimuli

The stimuli, presented on a screen situated outside the scanner, were reflected onto a mirror (of size 20 x 9 cm^2) placed inside the scanner. In each trial, two digits appeared simultaneously in white on a black background. Each presentation lasted 1000 ms, and there was an interval of 2000 ms before the subsequent trial.

The stimuli in the current experiment were the same as those used in Experiment 3a, i.e., Arabic digits (1 to 9) in font Arial. The only difference was in terms of the absolute size of the stimuli: each digit might appear in one of the nine different sizes (subtending approximately 3.8°, 4.7°, 5.2°, 6.0°, 6.6°, 7.4°, 8.2°, 8.8°, and 9.7°). The increase in absolute size was introduced for comfort of viewing in the MRI scanner and to make the physical sizes more discriminable. The mean ratio between adjacent sizes remained unchanged, i.e., 1.1. All other aspects with regard to the stimuli remained unchanged from Experiment 3a.

Twenty-five random blank trials were added to each task, making a total of 217 (i.e., 64 congruent trials + 64 neutral trials + 64 incongruent trials + 25 blank trials = 217). Same counter-balancing rules from Experiment 3a were applied.

3.2.2.3 Subjects

There were 18 right-handed subjects (11 males and 7 females), age ranged 21 to 38 (mean = 25.0 years, standard deviation = 4.0 years). They performed two tasks (numerical and physical comparison tasks), half of them participated in the numerical task first, and the other half the physical task first. All subjects had normal or corrected-to-normal eyesight.

3.2.2.4 Scanning Procedures and Imaging Data Processing

Whole-brain fMRI data were acquired on a 1.5 Tesla Magnetom VISION system (Siemens Sonata, Erlangen Germany). Functional images were obtained with a gradient echo-planar sequence using blood oxygenation level-dependent (BOLD) contrast, each comprising a full brain volume of 28 contiguous axial slices (3.5 mm thickness). Volumes were acquired continuously with a repetition time (TR) of 2.52 s. A total of 275 scans were acquired for each participant in 2 sessions (approximately 10 minutes each), with the first 6 volumes subsequently discarded to allow for T1 equilibration effects. During fMRI scanning, pupil diameter was recorded on-line by an infrared eye tracker. The data were analysed using SPM2 (Wellcome Department of Imaging Neuroscience; www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 6.1.0.450 Release 12.1. Individual scans were realigned, slice time-corrected, normalized to the MNI template with voxels of 2 x 2 x 2 mm^3 and spatially smoothed by an 8-mm FWHM Gaussian kernel using standard SPM methods.

Event-related activity for each voxel, for each condition and each subject was modelled using a canonical haemodynamic response function plus temporal and dispersion derivatives. Statistical parametric maps of the t-statistic (SPM{t}) were generated for each subject and the contrast images were further smoothed by an 8-mm FWHM Gaussian kernel.

At the second level random-effects analysis, a 2 x 4 ANOVA model was applied; the factors were task (numerical comparison task and physical comparison task) and trial type (congruent, neutral, incongruent, and error trials). Congruent and incongruent trials were modelled parametrically with respect to task-relevant distance and task-irrelevant distance, and neutral trials were modelled parametrically with respect to task-relevant distance. In the present study, congruent and neutral trials were classified as non-conflict trials, whereas, incongruent trials were conflict trials. T-contrasts were constructed to compare conflict and non-conflict trials, and error and correct trials in each task. Threshold significance was set at 0.001 uncorrected for multiple comparisons.

3.2.3 Results

3.2.3.1 Behavioural Data

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 2000 ms after the stimulus offset. Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were also removed.

ANOVAs were used to analyse mean error rates and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

A 2 x 3 repeated-measures ANOVA was conducted on mean error rates. The factors were task (numerical comparison task and physical comparison task) and congruity (congruent, neutral, and incongruent). The ANOVA revealed a significant main effect of congruity ($F_{(1,22)} = 38.10$, $p \leq 0.001$), a non-significant main effect of task ($F_{(1,17)} = 3.34$, *n.s.*), and a non-significant task x congruity interaction ($F_{(1,21)} < 1$, *n.s.*). Tests of within-subjects contrasts revealed a significant difference in error rates between incongruent and neutral trials ($F_{(1,17)} = 52.00$, $p < 0.001$), and a non-significant difference between neutral and congruent trials ($F_{(1,17)} < 1$, *n.s.*). During numerical comparisons, the mean error rates were 2.35%, 3.83%, and 10.77% for congruent, neutral, and incongruent trials respectively (see Figure 3.4). During physical comparisons, the mean error rates were 5.56%, 5.03%, and 12.34% for congruent, neutral, and incongruent trials respectively (see Figure 3.4).

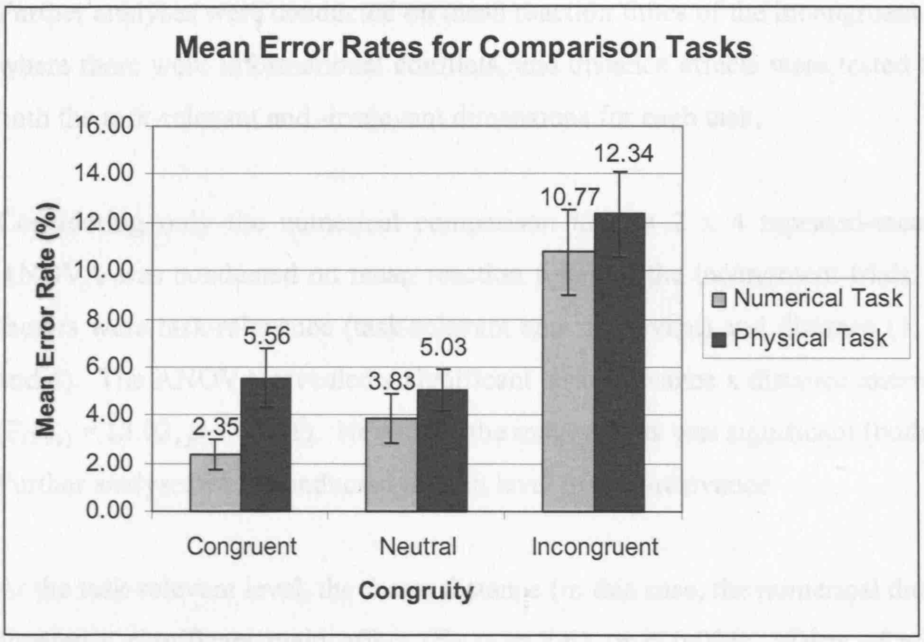


Figure 3.4 Mean error rates for numerical and physical comparison tasks (Experient 3b)

A 2 x 3 repeated-measures ANOVA was conducted on mean reaction times. The factors were task (numerical and physical comparison tasks) and congruent (congruent, neutral, and incongruent). The ANOVA revealed a significant main effect of task ($F_{(1,17)} = 16.55$, $p \leq 0.001$), a significant main effect of congruity ($F_{(2,34)} = 156.62$, $p < 0.001$), and a significant task x congruity interaction ($F_{(2,34)} = 8.07$, $p \leq 0.001$). The mean reaction times for numerical and physical comparison tasks were 617 ms and 570 ms respectively.

During numerical comparisons, tests of within-subjects contrasts revealed a significant difference between congruent and neutral trials ($F_{(1,17)} = 47.84$, $p < 0.001$) and between the latter and incongruent trials ($F_{(1,17)} = 81.60$, $p < 0.001$). The mean reaction times were 585 ms, 614 ms, and 652 ms respectively.

During physical comparisons, test of within-subjects contrasts revealed a significant difference between congruent and neutral trials ($F_{(1,17)} = 5.16$, $p < 0.050$) and between the latter and incongruent trials ($F_{(1,17)} = 56.16$, $p < 0.001$). The mean reaction times were 553 ms, 564 ms, and 593 ms respectively.

Further analyses were conducted on mean reaction times of the incongruent trials where there were informational conflicts, and distance effects were tested for in both the task-relevant and -irrelevant dimensions for each task.

Considering only the numerical comparison task, a 2 x 4 repeated-measures ANOVA was conducted on mean reaction times of the incongruent trials. The factors were task-relevance (task-relevant and -irrelevant) and distance (1, 2, 3, and 4). The ANOVA revealed a significant task-relevance x distance interaction ($F_{(3,51)} = 13.02, p < 0.001$). Neither of the main effects was significant (both *n.s.*). Further analyses were conducted at each level of task-relevance.

At the task-relevant level, the factor distance (in this case, the numerical distance) showed a significant main effect ($F_{(3,51)} = 8.61$, $p < 0.001$). Tests of within-subjects contrasts revealed a significant negative linear trend for this factor ($F_{(1,17)} = 19.20, p < 0.001$) (see Figure 3.5). No other trend was significant. At the task-irrelevant level, the factor distance (in this case, the physical distances) showed a significant main effect ($F_{(3,51)} = 4.78$, $p \leq 0.005$). Tests of within-subjects contrasts revealed a significant positive linear trend for this factor ($F_{(1,17)} = 7.51$, $p < 0.050$) (see Figure 3.5). No other trend was significant.

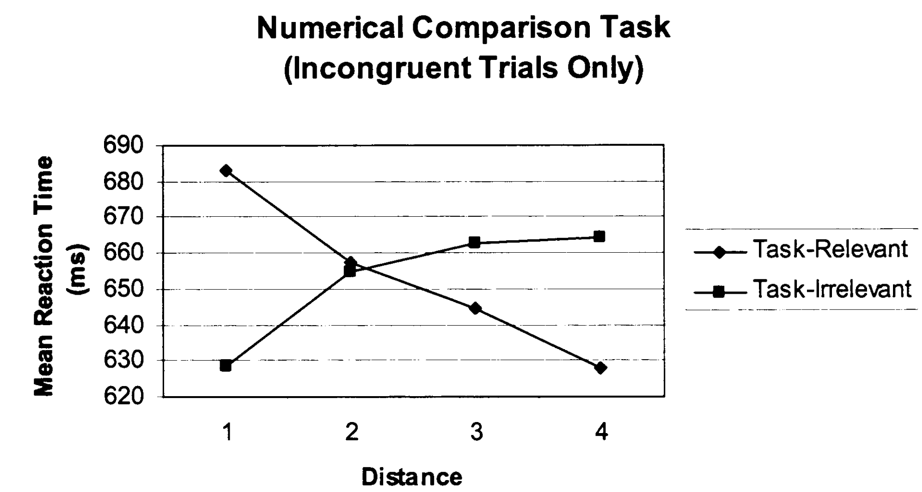


Figure 3.5 Mean reaction times (ms) for incongruent trials during numerical comparison task (Experiment 3b)

Considering only the physical comparison task, a 2 x 4 repeated-measures ANOVA was conducted on mean reaction times of the incongruent trials. The factors were task-relevance (task-relevant and -irrelevant) and distance (1, 2, 3, and 4). The ANOVA revealed a significant main effect of task ($F_{(1,17)} = 19.11, p < 0.001$), a significant main effect of distance ($F_{(3,51)} = 16.26, p < 0.001$), and a significant task-relevance x distance interaction ($F_{(3,51)} = 55.58, p < 0.001$). Further analyses were conducted at each level of task-relevance.

At the task-relevant level, the factor distance (in this case, the physical distance) showed a significant main effect ($F_{(2,33)} = 53.49, p < 0.001$). Tests of within-subjects contrasts revealed a significant negative linear trend ($F_{(1,17)} = 96.44, p < 0.001$) and a significant quadratic trend ($F_{(1,17)} = 8.98, p < 0.010$) for this factor (see Figure 3.6). At the task-irrelevant level, the factor distance (in this case, the numerical distances) showed a significant main effect ($F_{(3,51)} = 7.59, p < 0.001$). Tests of within-subjects contrasts revealed a significant positive linear trend ($F_{(1,17)} = 30.95, p < 0.001$) (see Figure 3.6). No other trend was significant.

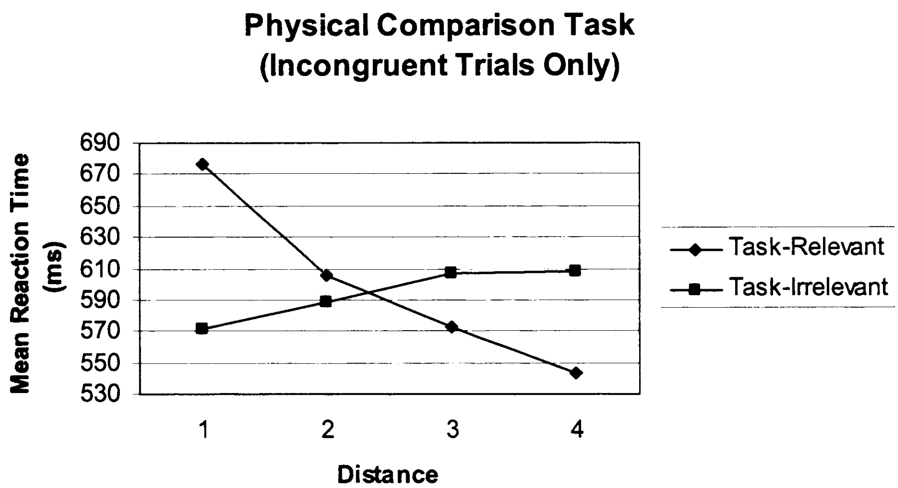


Figure 3.6 Mean reaction times (ms) for incongruent trials during physical comparison task (Experiment 3b)

In summary, during conflict trials, the relevant dimension showed a classic distance effect (indicated by a negative linear trend), whereas the irrelevant dimension showed a reversed distance effect (indicated by a positive trend), regardless of task.

3.2.3.2 Functional Imaging Data

Functional imaging data analysis at the first level allowed for neural responses associated with congruent, neutral, incongruent, and error trials to be modelled independently. Second level t-contrasts were constructed to test for brain regions associated with the parametric modulation of numerical distance and physical distance during neutral conditions. An F-contrast was then constructed to compare processing of the two dimensions. Analysis of the conflict trials allowed us to examine any task x distance interaction.

T-contrasts were used to identify regions involved in conflict and error trials. Conjunction analyses (by inclusive masking) were also performed to identify common regions for conflict and error processing across the two comparison tasks. Brain activations are summarised in Table 3.1, Table 3.2, and Table 3.3. Voxels reported are in Talairach coordinates.

In neutral trials, distance only varied parametrically in the task-relevant dimension, i.e., only numerical distance was manipulated in the numerical task, and only physical distance in the physical task. T-contrasts revealed no parietal region parametrically modulated by numerical distance or physical distance during these trials. However, the F-contrast comparing the processing of numerical distance and physical distance during neutral trials revealed several parietal regions which showed enhanced activation in processing numerical relative to physical distance including the right inferior parietal lobule [40 -39 42], right precuneus [22 -64 42], right inferior parietal lobule [32 -56 45], and left superior parietal lobule [-22 -66 46] (see Figure 3.7), as well as the bilateral inferior frontal gyri, right temporal and occipital regions (see Table 3.1). Small volume correction (SVC) searches (5 mm radius) were performed with reference to Pinel et al. (2004) and revealed that the right inferior parietal lobule [38 -41 41] also showed enhanced activity in numerical distance compared to physical distance. No enhanced activation was observed processing physical distance relative to numerical distance. Conjunction by inclusive masking across the two tasks revealed no commonly activated voxel.

Table 3.1 Brain activities associated with numerical and physical distance processing during neutral conditions (Experiment 3b)

Voxels	Z-Score	Talairach Coordinates			Brain Area
		x	y	z	
F-Contrast (Numerical Distance > Physical Distance)					
758	4.72	46	7	24	Right Frontal Lobe. Inferior Frontal Gyrus. White Matter.
169	3.86	-36	17	-3	Left Frontal Lobe. Inferior Frontal Gyrus. Grey Matter. Brodmann Area 47.
55	3.69	46	-77	17	Right Occipital Lobe. Middle Occipital Gyrus. White Matter.
501	3.68	40	-39	42	Right Parietal Lobe. Inferior Parietal Lobule. Grey Matter. Brodmann Area 40.
	3.53	22	-64	42	Right Parietal Lobe. Precuneus. White Matter.
	3.45	32	-56	45	Right Parietal Lobe. Inferior Parietal Lobule. White Matter.
95	3.45	50	-53	-9	Right Temporal Lobe. Sub-Gyral. White Matter.
	3.39	46	-61	-9	Right Occipital Lobe. Sub-Gyral. White Matter.
24	3.22	-22	-66	46	Left Parietal Lobe. Superior Parietal Lobule. Grey Matter. Brodmann Area 7.
23 (SVC)	3.67	38	-41	41	Right Inferior Parietal Lobule.

$p < 0.001$ uncorrected

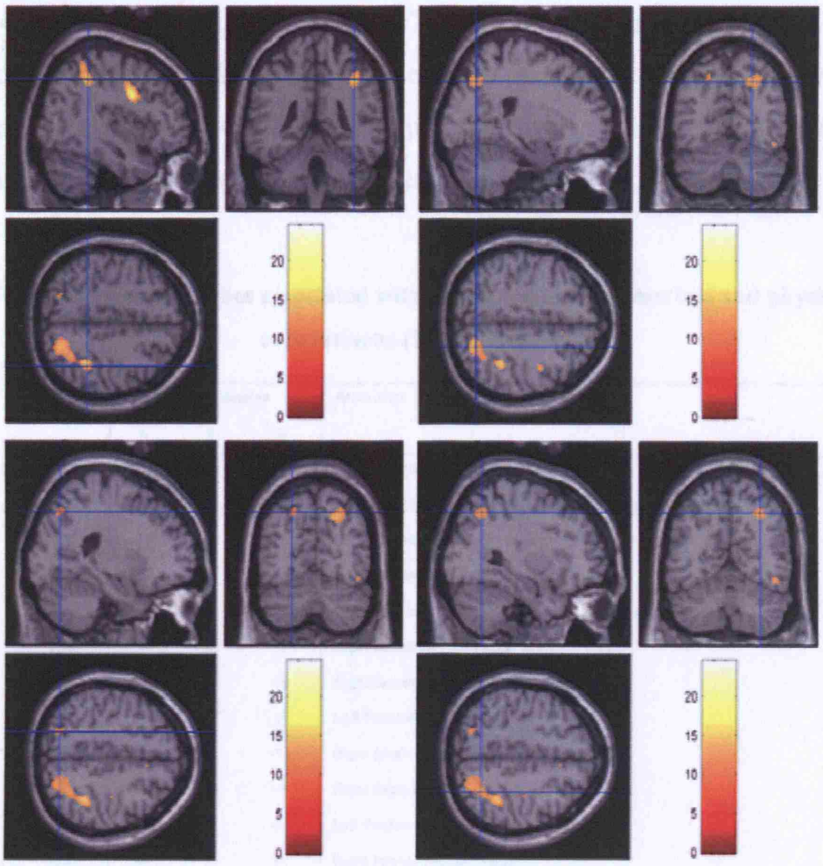


Figure 3.7 Parietal regions showing enhanced activation (p < 0.001 uncorrected) when processing numerical distance relative to physical distance (clockwise from top left: regions in the right inferior parietal lobule ([40 -39 42] and [32 -56 45]), right precuneus [22 -64 42], and left superior parietal lobule [-22 -66 46])

On the other hand, both numerical distance and physical distance were varied parametrically in incongruent trials. On the basis of autonomous processing of numerical magnitudes, the parietal regions which showed enhanced activation processing numerical relative to physical distance should not be affected by task requirement. Consistent with the prediction, the F-contrast which was constructed to test for a task x distance revealed no difference in parietal activation.

T-contrasts were constructed to test for differences in brain activity between conflict and non-conflict trials (see Table 3.2). In the numerical task, enhanced activation in conflict trials compared to non-conflict trials was observed in right inferior frontal [44 9 31] and middle frontal [36 52 -14] gyri, left fusiform gyrus [-44 -49 -13], right occipital lobe [34 -76 -1], and various other regions. In the physical task, enhanced activation in conflict trials was observed only in left inferior frontal gyrus [-40 30 8]. Conjunction by inclusive masking across the two tasks revealed no commonly activated voxel.

Table 3.2 Brain activities associated with conflicts during numerical and physical comparisons (Experiment 3b)

Voxels	Z-Score	Talairach Coordinates			Brain Area
		x	y	z	
T-Contrast (Numerical Task Conflict Trials > Numerical Task Non-Conflict Trials)					
224	4.32	32	-48	48	Right Superior Parietal Lobule. Brodmann Area 7.
112	4.08	-10	12	51	Left Superior Frontal Gyrus. Brodmann Area 6.
613	4.08	14	-5	11	Right Thalamus. Ventral Anterior Nucleus.
	3.93	18	6	9	Right Sub-lobar. Lentiform Nucleus. Putamen.
	3.60	8	6	0	Right Sub-lobar. Caudate. Caudate Head.
42	3.73	16	16	49	Right Superior Frontal Gyrus.
34	3.66	-22	-17	5	Left Sub-lobar. Extra-Nuclear.
16	3.43	36	52	-14	Right Middle Frontal Gyrus.
56	3.42	34	-76	-1	Right Occipital Lobe. Sub-Gyral.
14	3.28	-44	-49	-13	Left Fusiform Gyrus
28	3.25	44	9	31	Right Inferior Frontal Gyrus.
T-Contrast (Physical Task Conflict Trials > Physical Task Non-Conflict Trials)					
12	3.20	-40	30	8	Left Inferior Frontal Gyrus.

$p < 0.001$ uncorrected

T-contrasts were constructed to test for differences in brain activity between error and correct trials (see Table 3.3). Conjunction by inclusive masking (corrected for family wise error at 0.05) was performed across task and revealed enhanced activation in error trials compared to correct trials in bilateral inferior frontal gyri: left [-32 17 -11] and right [42 27 -6], and several regions along the bilateral superior temporal gyri (see Table 3.3).

Table 3.3 Brain activities associated with error processing during numerical and physical comparison tasks and conjunction (by inclusive masking) across them (Experiment 3b)

Voxels	Z-Score	Talairach Coordinates			Brain Area
		x	y	z	
T-Contrast (Numerical Task Error Trials > Numerical Task Correct Trials)					
2107	6.88	3	-42	11	Right Superior Temporal Gyrus.
	5.52	51	-26	-7	Right Temporal Sub-Gyral.
1200	5.32	-32	17	-11	Left Inferior Frontal Gyrus. Brodmann Area 47.
	3.99	-32	10	-29	Left Superior Frontal Gyrus
619	4.93	-51	-46	13	Left Superior Frontal Gyrus
3058	4.66	42	27	-6	Right Inferior Frontal Gyrus.
	4.56	55	20	12	Right Inferior Frontal Gyrus. Brodmann Area 45.
	4.56	34	16	-29	Right Superior Temporal Gyrus.
341	4.47	6	-15	3	Right Thalamus. Medial Dorsal Nucleus.
	4.08	-4	-15	3	Left Thalamus. Medial Dorsal Nucleus.
127	4.13	-18	-77	-25	Left Postertior Lobe. Uvula.
392	4.03	0	26	15	Inter-Hemispheric.
72	3.70	4	-22	-19	Right Brainstem. Pons.
166	3.65	-50	-2	31	Left Precentral Gyrus. Brodmann Area 6.
22	3.46	36	-5	-18	Right Temporal Sub-Gyral.
53	3.34	-46	13	20	Left Frontal Sub-Gyral.
27	3.25	-26	32	24	Left Frontal Sub-Gyral.
21	3.17	-59	-31	31	Left Inferior Parietal Lobule.
T-Contrast (Physical Task Error Trials > Physical Task Correct Trials)					
1807	5.83	53	-42	13	Right Superior Temporal Gyrus.
	5.08	50	-27	-5	Right Middle Temporal Gyrus.
1837	5.00	53	23	1	Right Inferior Frontal Gyrus.
	4.43	32	19	-9	Right Inferior Frontal Gyrus. Brodmann 47.
909	4.89	-34	17	-11	Left Inferior Frontal Gyrus. Brodmann 47.
335	4.15	-16	-73	-25	Left Posterior Lobe. Uvula.
149	4.04	4	-35	-35	Right Brainstem. Medulla.
Conjunction by Masking Across Tasks (Error Trials > Correct Trials), corrected for family wise error at 0.050					
715	6.88	53	-42	11	Right Superior Temporal Gyrus.
72	5.52	51	-26	-7	Right Superior Temporal Sub-Gyral.
181	5.32	-32	17	-11	Left Inferior Frontal Gyrus. Brodmann Area 47.
37	4.93	-51	-46	13	Left Superior Temporal Gyrus.
32	4.66	42	27	-6	Right Inferior Frontal Gyrus.

$p < 0.001$ uncorrected

3.2.4 Discussion

3.2.4.1 The Stroop Effect, Distance Effects, and Autonomous Processing of Numerical Magnitude Information

The classical Stroop effect was observed. Reaction time analysis revealed significant interference and facilitation in both numerical and physical tasks, consistent with the number Stroop literature (e.g., Girelli et al., 2000). Thus, subjects experienced difficulties in ignoring the irrelevant information from the irrelevant dimensions, regardless of task. Facilitation was reflected by reaction times but not error rates, consistent with the general finding in the Stroop literature that facilitation is virtually always substantially smaller than interference (see review, MacLeod, 1991).

A classic distance effect (indicated by a negative linear trend) was observed in both dimensions when task-relevant, i.e., numerical distance effect in the numerical task (replicating previous research findings, e.g., Banks et al., 1976; Duncan & MacFarland, 1980; Fias et al., 2003; Foltz et al., 1984; Hinrichs et al., 1981; Moyer & Landauer, 1967; Parkman, 1971; Pinel et al., 2001, 2004; Sekuler & Mierkiewicz, 1977) and physical distance effect in the physical task (replicating finding of Experiment 2 of the current thesis).

By increasing the absolute sizes of the stimuli (so that adjacent sizes differed by between 0.5° and 0.9°, as opposed to between 0.2° and 0.5° in Experiment 3a), the physical sizes of the stimuli were disambiguated, and a reversed distance effect was observed in both dimensions when task-irrelevant, i.e., a reversed numerical distance effect in the physical task (replicating findings of Girelli et al., 2000; Henik & Tzelgov, 1982; see also Experiment 2 of the current thesis), as well as a reversed physical distance effect in the numerical task (replicating findings of Experiments 2 and 3a of the present thesis).

Distance effects were used as indicators for information processing in the present experiment, and the observed reversed distance effects under task-irrelevant conditions indicate that numerical distance and physical distance were processed in a refined autonomous fashion, similar to information processing under task-relevant conditions. The current findings therefore provide strong evidence for refined autonomous processing of numerical magnitude information.

When appropriate stimuli (in terms of range of numbers and their physical size) were chosen, the reversed numerical distance effect first observed by Henik and Tzelgov (1982) was reliably replicated (Experiments 2 and 3b of the present thesis). The effect has proved to be a robust indicator of semantic processing and a more sensitive measure than the Stroop effect itself in experiments where interference was not consistently elicited.

By incorporating the full range of single digits (1-9) and 9 physical sizes, 4 levels of distance (instead of only 3 in Experiment 2) were constructed and used in the present experiment, and the distance effects currently observed (see Figure 3.5 and Figure 3.6) were more prominent than those in Experiment 2 (see Figure 2.17 and Figure 2.18).

The fact that the reversed numerical distance effect was not always observed or even examined in previous experiments could be attributed to poor experimental designs. Two levels of numerical distance were often employed, giving rise to distant and close trials, but if these conditions were similar (i.e., the two numerical distances were not highly discriminable), it would be difficult for a distance effect to emerge, especially under task-irrelevant condition. To illustrate this point further, one could examine Figure 3.6; if numerical distances 3 and 4 were chosen, a reversed numerical distance effect would be unlikely to emerge). Moreover, the use of two levels of distance meant that only a mere difference could be established between the distant and the close conditions, but not a linear trend per se. The present experiment (and Experiments 2 and 3a) employed more levels of distance, allowing polynomial contrasts to be specified in the analyses, thus linear (and any other higher order) trends could be established.

As proposed in the Introduction of Experiment 3a (Section 3.1.1), the reversed distance effects may be explained in terms of amount of interference. Information that requires little effort to process under task-relevant conditions (e.g., the salient difference between two numbers with a large physical distance) would be harder to ignore under task-irrelevant conditions (e.g., during numerical comparisons). In contrast, information that requires more effort to process under task-relevant conditions (e.g., numerical comparison of two numerically close numbers) would exert little interference when such information is to be ignored under task-irrelevant condition (e.g., during physical comparisons). For example, selecting the numerically larger item would be slower for 3 5 than 3 5 since it is quicker to resolve the physical size difference in the first case. The current findings are consistent with those of Pansky and Algom (1999) who found, in number Stroop experiment, that when the task-irrelevant dimension was more discriminable, a sizeable Stroop effect affected performance on the task-relevant dimension, but when it was less discriminable, the Stroop effect was considerably weaker.

3.2.4.2 Parietal Activities and Numerical Information Processing

Neuroimaging data revealed that when the task-irrelevant dimensions were kept constant (i.e., in neutral trials), the parietal lobes (regions in right inferior parietal lobule, right precuneus and left superior parietal lobule) showed enhanced activation when processing numerical distance compared to physical distance. This extends the findings of Pinel et al. (2004) who observed enhanced activation in parietal lobes (regions in bilateral inferior parietal foci and left intraparietal sulcus) during numerical comparisons compared to physical comparisons, implying a processing difference between numerical and physical dimensions.

The enhanced activation during numerical processing observed in the parietal regions when processing numerical distance relative to physical distance is consistent with (though does not demonstrate) the distinction made by Zorzi and Butterworth (1999) between numerical magnitudes which are conceptualised as "discrete numerosities" and physical sizes which are represented in an analogue (or continuous) fashion, and suggests that comparative judgements on discrete numerosity representations evoked by numbers call for higher processing requirements compared to those on analogue representations evoked by physical sizes. The present finding is also in agreement with Castelli et al.'s (2006) finding that the intra-parietal sulcus showed enhanced activation when viewing discrete stimuli compared to continuous stimuli.

As predicted, the parietal regions (in right inferior parietal lobule, right precuneus and left superior parietal lobule) which showed enhanced activation processing numerical relative to physical distance were not affected by task requirements in conflict situations. In other words, these parietal regions were equally active whether or not required by the task to process numerical magnitudes. The lack of difference in parietal activation level across numerical and physical comparison tasks during conflict situations provides a strong piece of evidence for autonomous processing of numerical magnitude.

Interestingly, no evidence was found to suggest that parietal activation was parametrically modulated by either numerical or physical distance. This appears inconsistent with numerical distance modulated parietal regions identified by Pinel and colleagues (Pinel et al., 2001, 2004). However, a closer look at their findings suggests that the numerical distance modulated parietal activation might not be as robust as the authors claimed. In Pinel et al.'s (2001) study, numerical distances were not modelled parametrically in the analysis. The reported distance modulated parietal activation was in fact a main effect of distance (across three levels) rather than a significant linear decrease in activation with increasing numerical distance. Furthermore, when masking the main effect of distance for close > medium and medium > far, the only surviving significant region was in the precuneus. In addition, the reported numerical distance modulated parietal regions in the bilateral intraparietal sulci and the right precuneus were from a single-subject analysis. Of the four subjects tested, only two showed similar effects; one failed to show any strong correlation between brain activation and numerical distance. The inconsistent findings cast doubts on whether parietal activation is truly modulated by numerical distance.

One could argue that the current paradigm is more complex than the one used by Pinel et al. (2001), and this could have meant that any potential numerical distance modulated parietal activation was less likely to emerge. Pinel et al. (2001) employed a comparison task to a fixed reference, in which subjects had to judge whether a visually presented number was smaller or larger than the reference (65). Such a task is probably easier than the current paradigm which involves comparing multidimensional stimuli. Since Pinel et al.'s (2001) stimuli only varied in one dimension (numerical magnitude), there was neither task-irrelevant information nor conflict to influence subjects' judgements, and the subjects were likely to perform the task without much difficulty.

The current experiment and Pinel et al.'s (2004) experiment employed the same comparison Stroop paradigm. Both used 4 levels of numerical distance: 1, 2, 6, and 7 in Pinel et al.'s (2004) study and 1, 2, 3, and 4 in the current study. However, Pinel et al. (2004) grouped distances 1 and 2 as close and distances 6 and 7 as far. Thus, the distance related parietal activation was merely a difference between the two levels; as in Pinel et al.'s (2001) experiment, a linear decrease in parietal activation with increasing numerical distance could not be established due to the limited levels of distance used in the analysis. The present experiment tested precisely the latter (a linear change with increasing distance) by explicitly modelling numerical distance (and physical distance) parametrically in the analysis, but no evidence was found to support a linear decrease (or increase) with increasing numerical (or physical) distance.

3.2.4.3 Brain Regions Associated with Conflict Resolution and Error Commission

When confronted with a conflict, subjects had to inhibit the information from the task-irrelevant dimension in order to perform the task correctly. The enhanced activity in the right inferior frontal gyrus in conflict trials compared to non-conflict trials during numerical comparisons is consistent with findings in go/no-go and stop-signal tasks (e.g., Bunge et al., 2002; Garavan et al., 1999; Konishi et al., 1998, 1999; Menon et al., 2001; Rubia et al., 1999). In Durston et al.'s (2002) go/no-go task, activity associated with successful response inhibition extended to right middle frontal gyrus. In the current experiment, enhanced activity in this area was also observed in conflict trials during numerical comparisons.

The left fusiform gyrus showed enhanced activity in conflict trials during numerical comparisons. Such activity does not only reflect the recognition of the visual shapes of the numbers, but also suggests that the computation over the perceptual properties of the numbers interacted with the semantic processing of the task-relevant dimension of the stimuli, i.e., the numerical magnitudes. This suggestion is further supported by the enhanced activation in the right occipital lobe – an area involved in visual processing – in conflict trials during numerical comparisons.

During physical comparisons, enhanced activity in the left inferior frontal gyrus was observed in conflict trials, but there was no evidence to suggest that differential activity occurred in the visual cortex, supporting the idea that the interference comes from higher cognitive processing of the numbers.

There was no evidence to support the proposed involvement of the ACC in conflict resolution. Small volume correction searches (5 mm radius) were performed with reference to ACC co-ordinates identified by Barch, Braver, Nystrom, Forman, Noll, and Cohen (1997), Bush et al. (1999), Menon et al. (2001), and van Veen, Cohen, Botvinick, Stenger, and Carter (2001), but no significant voxel was found.

The absence of enhanced ACC activation during conflict trials compared to non-conflict trials can be explained in terms of the sequential events in a comparative judgement. In the present comparison paradigm, interference (between numerical magnitude and physical size) is separated from response competition (between left and right hand key presses); thus it is likely that conflict resolution takes place prior to response selection. Furthermore, all trial-types present the same amount of response competition (between the two key presses). Since ACC activation reflects conflict at the response level (Bunge et al., 2002; Milham, Banich, & Barad, 2003; Milham, Banich, Webb, Barad, Cohen, Wszalek, & Kramer, 2001; Nelson, Reuter-Lorenz, Sylvester, Jonides, & Smith, 2003; van Veen et al., 2001; Weissman, Giesbrecht, Song, Mangun, & Woldorff, 2003), the absence of enhanced ACC activation during conflict trials relative to non-conflict trials was unsurprising.

There was also no evidence to support the involvement of the ACC in error commission. Small volume correction searches (5 mm radius) were performed with reference to ACC co-ordinates identified by Menon et al. (2001) who reported its associated activity with error processing, but no significant voxel was found. Since error-associated ACC activation reflects merely the detection of a post-response conflict, the absence of such activation is not surprising given that conflict resolution in the current comparative judgements is likely to take place prior to response selection.

The present findings provide evidence that bilateral inferior frontal gyri are involved in conflict resolution. The fact that these areas also showed enhanced activation during error trials can be explained in terms of the disproportionately high percentage of errors occurring in incongruent (conflict) trials.

3.2.5 Summary

The present experiment has provided evidence for autonomous processing of numerical magnitude at both behavioural level (indicated by the reversed numerical distance effect) and neuronal level. Although there was no evidence to suggest that parietal activity was parametrically modulated by numerical distance, enhanced activation was observed in several parietal regions (in right inferior parietal lobule, right precuneus and left superior parietal lobule) when processing numerical magnitudes relative to physical sizes, and such an enhancement was not affected by task requirement in conflict situations. In other words, these parietal regions were equally active whether or not required by the task to process numerical magnitudes. The lack of difference in parietal activation level across numerical and physical comparison tasks during conflict situations provides a strong piece of evidence for autonomous processing of numerical magnitude.

4 Comparing Numerical Magnitudes and the Effects of Familiarity

4.1 Introduction

The current section discusses the effects of familiarity on the Stroop phenomenon. The traditional colour-word Stroop effect that words interfere with colour naming may be viewed as a result of our extensive reading experience. Support of this view comes from the developmental trend of the Stroop effect; the interference exerted by task-irrelevant numerical magnitudes on physical comparisons "was absent in first-grade children's performance, emerged in the third grade, and was highly significant in the fifth grade" (Girelli et al., 2000). These observations suggest that a firm understanding of numerical magnitude from practice is necessary for the semantic aspect of numbers to interfere with physical size comparisons. An obvious and theoretically critical question which follows is: does the Stroop effect change as a consequence of practice at the task? Intuitively, extended practice with the Stroop task should lead to a reduction of the Stroop effect as subjects develop a strategy for coping more successfully with the task. This finding has indeed been observed (e.g., Effler, 1978; Ogura, 1980), but not always (e.g., Harbeson, Kennedy, & Bittner, 1982; Shor, Hatch, Hudson, Landrigan, & Shaffer, 1972; White, 1978).

Some authors have suggested that the reduction of the Stroop effect may be dependent on the mode of response (e.g., Flowers & Stoup, 1977; Nielsen, 1975; Roe, Wilsoncraft, & Griffiths, 1980). In particular, Nielsen (1975) found an initially greater Stroop effect with a vocal than a manual response, but then observed a greater decline of the effect over practice with the latter. Roe et al. (1980) confirmed that manual responding is affected more quickly by practice, which is interesting given that manual responding is faster to begin with and therefore close to a "performance floor".

Research has also suggested a good deal of specificity to practice effects. Reisberg, Baron, and Kemler (1980), in an enumeration Stroop task (where subjects had to enumerate the number of items presented while ignoring the identity of the stimuli), observed a decline in the Stroop effect but in a very specific way. Subjects who had practiced ignoring the digits 2 and 4 showed no benefit in an enumeration test where they had to ignore digits 1 and 3 or the words "to" and "for". However, there was good transfer when the written verbal numerals "two" and "four" were to be ignored. These findings provide strong evidence for practice effects and semantic transfer in Stroop performance.

Reasoning that a novel (unfamiliar) orientation would slow down word processing, Liu (1973), in a modified version of the colour-word Stroop task, required subjects to name the ink colours while the orientation of the stimuli (colour words) varied. The Stroop effect was found to have reduced compared to upright (familiar) orientation. However, Dunbar and MacLeod (1984) failed to replicate Liu's (1973) results. Instead, they observed equivalent interference regardless of the word's orientation, suggesting that previous practice could not be the critical factor in determining the pattern of the Stroop effect.

Regan (1977) compared a well-practised task to a newly acquired task. English-speaking subjects had to learn the names of Armenian letters, then to name the small letters while ignoring the large letter formed by the small letters (both small and large letters could either be English or Armenian, see Navon's (1977) procedure). The Stroop effect observed was equivalent for the very familiar, well-practised set (English) and for the newly acquired set (Armenian). Such a finding suggests that autonomous processing of the task-irrelevant dimension may not necessarily require extensive practice. This challenges the hypothesis that the Stroop phenomenon is a direct consequence of differential practice. However, no alternative explanation was provided.

It is clear that the effects of practice warrant further investigation. Experiment 4 examines such effects with stimuli which people have daily encounter with – coins. On the basis that the Stroop effect is a direct consequence of differential practice, extended practice with the task should reduce the effect. It follows that, a conflict which is familiar (well-practised) would induce a weaker Stroop effect than one which is unfamiliar (novel). In the present experiment, images of the

British coins were used to investigate the effects of familiarity on the Stroop effect. The images were modified to create both familiar and unfamiliar stimuli. Since the value of British coins increases in a non-systematic way with their physical size, some coin pairs pose a conflict with respect to the two dimensions (conceptual value and physical size), for example, a 5p coin is physically smaller than a 2p coin although the former has a larger value. The present experiment was a variant of the comparison Stroop paradigm; subjects had to judge the larger of a coin image pair in either conceptual value or physical size.

Based on the assumption that the Stroop effect is a direct consequence of practice, it was predicted that extensive experience from daily encounter with the British coins would lead to the development of a coping strategy with the familiar conflicts (e.g., a small 5p and a large 2p), and hence a lesser degree in the Stroop effect when confronting with these conflicts compared to unfamiliar ones (e.g., a small £2 and a large £1).

4.2 Experiment 4: Comparing Conceptual Values and Physical Sizes of British Coin Images

4.2.1 Methods

4.2.1.1 Tasks

The two tasks were conceptual value comparison and physical size comparison. In the former, subjects had to choose the coin which was conceptually larger (i.e., the one with the larger value), and in the latter, subjects had to choose the coin which was physically larger. Subjects had to make a manual response (by left or right clicking the mouse) in each trial to indicate the side of the larger relevant attribute. Their reaction times and responses were recorded.

The tasks were computer-based. A program written in Visual Basic 6.0 was used. It obtained inputs from Microsoft Excel files and recorded outputs (reaction times and key responses) onto text files.

4.2.1.2 Stimuli

The stimuli were images of the current British coins, modified using Adobe Photoshop 6.0. The images, originally obtained from The British Royal Mint website (http://www.royalmint.com/), were modified to monochrome, adjusting for brightness and contrast. This was done to prevent subjects from using colour cues to solve the comparisons. Only the tail sides, where the values of the coins are present, were used. Images of all the coins were used: 1p, 2p, 5p, 10p, 20p, 50p, £1, and £2 (see Table 4.1 for actual dimensions), but not all coin pairs were used. In order to select the coin pairs to be used, the coins were ranked in their actual sizes as given in The British Royal Mint website (http://www.royalmint.com/), see Table 4.2. Coins which are adjacent to one another along the physical dimension were not used because some of the neighbouring coins are too similar in physical size, and hence physical size comparisons between them would be difficult. Thus, the 21 pairs used were: 5p-20p, 5p-£1, 5p-10p, 5p-2p, 5p-50p, 5p-£2, 1p-£1, 1p-10p, 1p-2p, 1p-50p, 1p-£2, 20p-10p, 20p-2p, 20p-50p, 20p-£2, £1-2p, £1-50p, £1-£2, 10p-50p, 10p-£2, and 2p-£2.

Table 4.1 Diameters (mm) of British coins in descending coin value order (Experiment 4)

Coin	Diameter (mm)
£2	28.40
£1	22.50
50p	27.30
20p	21.40
10p	24.50
5p	18.00
2p	25.90
1p	20.03

Table 4.2 Diameters (mm) of British coins in ascending diameter order (Experiment 4)

Coin	Diameter (mm)
5p	18.00
1p	20.03
20p	21.40
£1	22.50
10p	24.50
2p	25.90
50p	27.30
£2	28.40

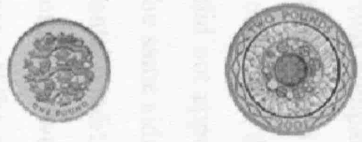
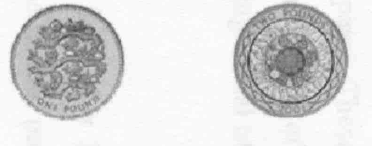
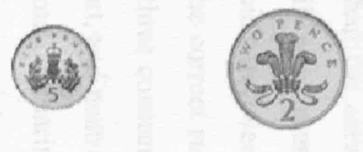
In each trial, an 800 ms fixation point (x) appeared on a white background in the middle of a 14-inch screen, followed by the presentation of the stimuli, a pair of coin images. The stimuli remained on the screen until the subject had made a key response. Then, there was an interval of 1000 ms before the subsequent trial.

There were three experimental conditions: congruent, when the conceptually larger coin was physically larger; incongruent, when the numerically larger coin was physically smaller; and neutral, when the coins were of the same physical size in the conceptual comparison task or when the same coins appeared in different sizes in the physical comparison task.

The experiment aimed to investigate the influence of experience on the Stroop effect, so an additional factor, familiarity, was introduced. Familiarity is defined in terms of the physical size ratio between the coin images. For example, the pair 5p and 1p in their normal ratio was classified as a familiar incongruent pair (since the 5p coin which has a larger value is physically smaller then the 1p coin). To create unfamiliar pairs, the physical size ratio between each coin pair used in the experiment was reversed, so for the 5p and 1p pair, the latter would now appear physically smaller than the former, thus an unfamiliar congruent pair.

Since two different coins can never have the same diameter, all neutral pairs are, by definition, unfamiliar. In the conceptual comparison task, coin images in a neutral pair were of the same physical size. These neutral pairs were created by using the mean of the diameters of the coins in a given pair. On the other hand, a neutral pair in the physical comparison task consisted of the same coin in two different sizes. To create these neutral pairs, the average difference of the diameters between the coins in each of the 21 pairs was calculated. This value was then divided by 2, giving a constant of 2.72 mm. Each of the 8 coin images was then resized to produce 2 images for each neutral pair, so that the diameter of the physically smaller one (d_s) would be given by ($d_s = d_o - 2.72$) and that the diameter of the physically larger one (d_l) would be given by ($d_l = d_n + 2.72$), where (d_o) refers to the diameter of the original coin. Sample stimuli for the present experiment are shown in Table 4.3.

Table 4.3 Sample stimuli (Experiment 4)

		Physical Size Ratio between Coin Images	
		Familiar	**Unfamiliar**
Congruent			
Neutral	Conceptual Task	-	
	Physical Task	-	
Incongruent			

The 21 pairs mentioned above, in their original, unresized ratios, were familiar pairs, of which 16 were congruent and 5 were incongruent. When the sizes of these coins in each pair were reversed to form unfamiliar pairs, there were 5 congruent and 16 incongruent pairs. These 21 pairs were also resized to form 21 neutral pairs, each of which consisted of two identical size coins. Each pair would appear twice, thus making a total of (21 + 21 + 21) x 2 = 126 trials in the numerical comparison task.

The same (21 x 2) congruent and (21 x 2) incongruent pairs were used in the physical comparison task. The neutral pairs were formed by presenting two of the same coin in different physical sizes. All 8 coins were used, thus there were 8 possible pairs. Each of these 8 pairs was presented 5 times, giving a total of 40 neutral pairs. Thus, the grand total number of trials in the physical comparison task was (42 + 42 + 40) = 124 trials.

In each of the two comparison tasks, half of the trials had the correct answer appearing on the left, and the rest on the right. Stimuli were presented in a pseudorandom order with the following constraints to avoid carryover effects: (1) the same coin did not appear in consecutive trials, (2) the correct response did not appear on the same side (left or right) for more than three consecutive trials, (3) the experimental condition (i.e., congruent, incongruent, and neutral) was not the same for more than two consecutive trials, (4) the same familiarity (familiar or unfamiliar) was not the same for more than four consecutive trials.

4.2.1.3 Subjects

The subjects were 14 undergraduate students (7 males and 7 females) at University College London, aged from 19 to 23 (mean 20.6 years, standard deviation = 1.2 years). All subjects had always resided in the United Kingdom prior to the experiment and had therefore had extensive experience with British coins.

4.2.2 Results

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 1000 ms after the stimulus offset. Error rates were generally low, but three subjects who made more than 10% error in the conceptual task were removed from the analyses. In the remaining subjects (N=11), error rates were significantly higher in the conceptual task than in the physical task ($t_{(10)} = 3.76$, $p < 0.005$) – the mean error rates were 5.69% and 2.77% respectively. Error trials were removed from further analyses. Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were also removed from further analyses.

The experimental design includes 3 factors: task (conceptual and physical), congruity (congruent, neutral, and incongruent), and familiarity (familiar and unfamiliar). Since all neutral trials are, by definition, unfamiliar, they would not form an appropriate baseline against congruent and incongruent trials, hence were removed from the analyses.

A 2 x 2 x 2 repeated-measures design ANOVA was conducted on mean reaction times. The factors were: task (conceptual and physical), congruity (congruent and incongruent), and familiarity (familiar and unfamiliar). The ANOVA revealed a significant main effect of task ($F_{(1,10)} = 62.73$, $p < 0.001$) – physical comparisons (415 ms) were significantly faster than conceptual comparisons (756 ms). There was also a significant main effect of congruity ($F_{(1,10)} = 8.10$, $p < 0.050$) – congruent trials (568 ms) were responded to significantly faster than incongruent trials (604 ms). The main effect of familiarity was not significant ($F_{(1,10)} < 1$, *n.s.*). There were two significant interactions: task x familiarity ($F_{(1,10)} = 6.10$, $p < 0.050$) and task x congruity x familiarity ($F_{(1,10)} = 8.44$, $p < 0.050$). No other significant interaction was found.

In the conceptual comparison task, there was a significant main effect of congruity ($F_{(1,10)} = 6.30$, $p < 0.050$) – congruent trials (727 ms) were responded to significantly faster than incongruent trials (786 ms). The main effect of familiarity was not significant ($F_{(1,10)} = 2.58$, *n.s.*). However, there was a significant congruity x familiarity interaction ($F_{(1,10)} = 6.35$, $p < 0.050$). Paired samples t-tests revealed a significant difference in mean reaction times between familiar congruent trials (696 ms) and unfamiliar congruent trials (757 ms) ($t_{(10)} = 4.22$, $p < 0.005$), but the difference in mean reaction times between familiar incongruent trials and unfamiliar incongruent trials was not significant ($t < 1$, *n.s.*). Figure 4.1 shows the mean reaction time pattern for the conceptual comparison task.

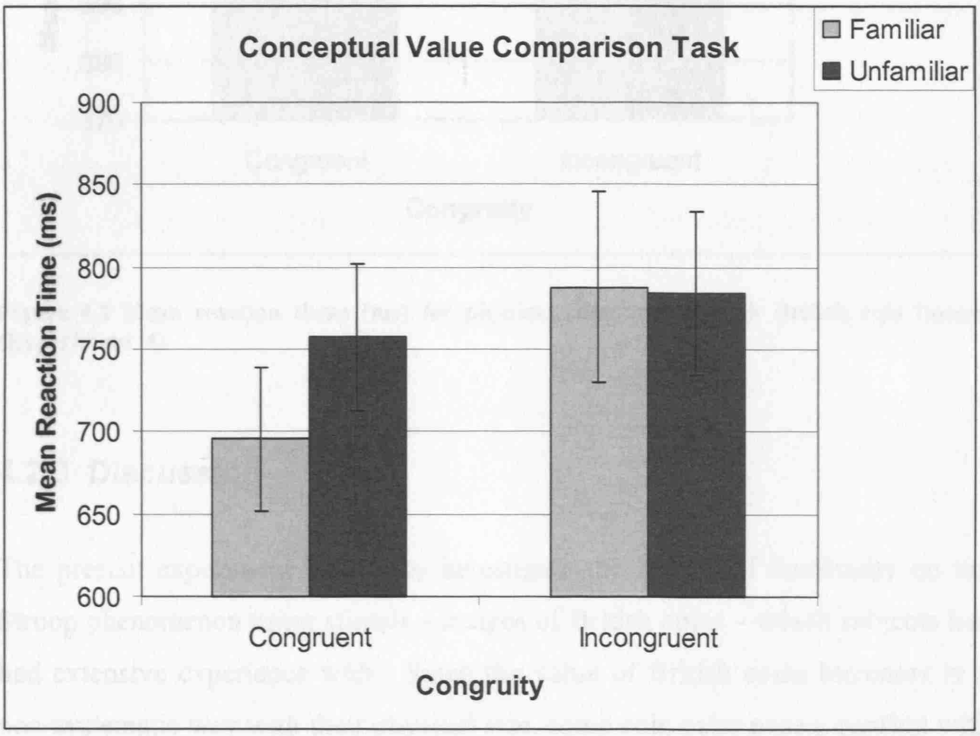


Figure 4.1 Mean reaction times (ms) for conceptual comparisons with British coin images (Experiment 4)

In the physical comparison task, the main effect of congruity was significant ($F_{(1,10)} = 6.02$, $p < 0.050$) – congruent trials (409 ms) were responded to significantly faster than incongruent trials (422 ms). The main effect of familiarity and the congruity x familiarity interaction were not significant (all *n.s.*). Figure 4.2 shows the mean reaction time pattern for the physical comparison task.

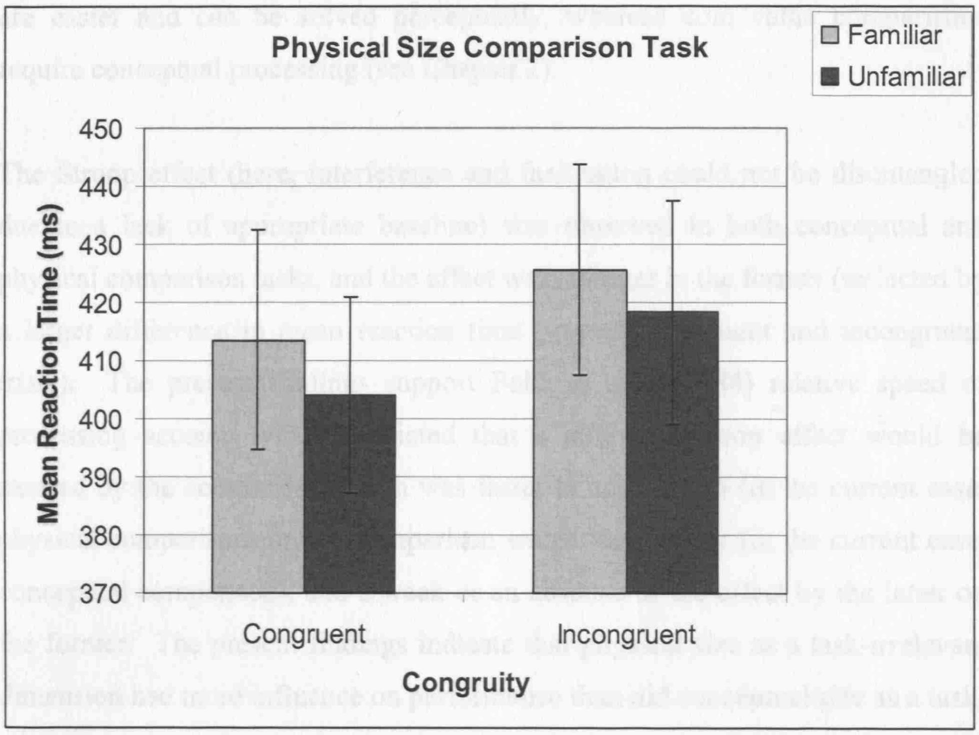


Figure 4.2 Mean reaction times (ms) for physical comparisons with British coin images (Experiment 4)

4.2.3 Discussion

The present experiment set out to investigate the effects of familiarity on the Stroop phenomenon using stimuli – images of British coins – which subjects had had extensive experience with. Since the value of British coins increases in a non-systematic way with their physical size, some coin pairs pose a conflict with respect to the two dimensions (conceptual value and physical size), for example, a 5p coin is physically smaller than a 2p coin although the former has a larger value. The experiment examined whether extensive experience from daily encounter of these stimuli would alter the pattern of the Stroop effect.

The conceptual comparison task was found to be more difficult than the physical comparison task, evidenced by the significantly higher mean error rate in the latter. It is therefore not surprising that physical comparisons were executed significantly faster than conceptual comparisons. These findings are consistent with those of Experiments 1, 2, 3a and 3b, suggesting that physical comparisons are easier and can be solved perceptually, whereas coin value comparisons require conceptual processing (see Chapter 2).

The Stroop effect (here, interference and facilitation could not be disentangled due to a lack of appropriate baseline) was observed in both conceptual and physical comparison tasks, and the effect was stronger in the former (reflected by a larger difference in mean reaction time between congruent and incongruent trials). The present findings support Foltz et al.'s (1984) relative speed of processing account which predicted that a stronger Stroop effect would be exerted by the comparison which was faster to accomplish (in the current case, physical comparison) on the comparison which was slower (in the current case, conceptual comparison), and a weak or an absence of the effect by the latter on the former. The present findings indicate that physical size as a task-irrelevant dimension had more influence on performance than did conceptual size as a task-irrelevant dimension, mirroring Henik and Tzelgov's (1982) finding with numbers that "Physical size as an irrelevant dimension had more influence on performance than did semantic size as an irrelevant dimension."

Due to the complexities of the stimuli used in the present experiment (non-systematic changes in physical size with increasing conceptual coin values), conceptual and physical distance effects were not examined. Thus, although the Stroop effect was observed in both conceptual and physical tasks, whether or not the corresponding task-irrelevant dimension (physical distance and conceptual distance respectively) was processed in a refined fashion could not be determined.

The present experiment examined the effects of familiarity on the Stroop phenomenon. Although the stronger Stroop effect observed during conceptual comparisons was consistent with Foltz et al.'s (1984) relative speed of processing account, no explanation could be offered by this account to explain the interactions between familiarity and other factors, namely task and congruity. When confronted with conflicts in both conceptual and physical tasks, familiarity with coin ratios did not aid comparison judgements. The current finding therefore provides no support to the prediction that that extensive experience with the British coins would allow subjects to develop a coping strategy with familiar conflicts (e.g., a small 5p coin and a large 1p coin). The finding appears consistent with Regan's (1977) in that the pattern of the Stroop effect did not depend on extensive training, suggesting that once established, the effect would be robust and resistant to changes by extensive practice at the task.

However, the results were less straightforward when familiar trials during conceptual comparisons were considered. An intriguing finding was observed such that when the physical size was congruent to the coin value, familiar pairs had a significant advantage over unfamiliar pairs. The equivalent advantage brought about by the familiar trials was not observed during physical comparisons.

Taken together, familiarity of the stimulus ratios had differential effects on conflict (incongruent) and non-conflict (congruent) trials during conceptual comparisons. There was no evidence to suggest that familiarity would aid conflict resolution, however it did benefit non-conflict trials. One possible explanation for the differential familiarity effects is that there are fewer "naturally-occurring" (familiar) conflict pairs than non-conflict pairs of coins – of the 21 pairs used in the present experiment, only 5 contained a conflict, and this might have led to better-practised congruent trials relative to incongruent trials. Within Cohen et al.'s (1990) framework, it would mean that familiar congruent trials had established a stronger pathway due to their more frequent exposures compared to the familiar incongruent trials, and as a result, familiarity aided congruent trials. On the other hand, the relative infrequent exposures of familiar incongruent trials meant a relatively weaker pathway, thus familiarity provided no benefit over incongruent trials. In other words, the extent of practice was different with congruent and incongruent stimulus pairs (more frequent for the former).

The above suggestion however does not explain why the equivalent familiarity effect was not observed during physical comparisons. The finding that familiarity only aided conceptual comparisons suggests that there might be an interaction between familiarity and stimulus discriminability. Since physical comparisons can be solved perceptually, whereas coin value comparisons require conceptual processing, it would be reasonable to assume that the physical dimension is more easily discriminated than the conceptual dimension. The higher discriminability of the congruent trials combined with their familiarity might have contributed to the reduction in mean reaction time during conceptual comparisons. On the other hand, familiarity had no effect on the Stroop phenomenon when the task-irrelevant dimension – conceptual value – was relatively low in discriminability.

Note that the above explanations are speculative and further research is necessary to examine these suggestions.

The use of coin images in the present experiment was an attempt to investigate a type of stimulus which had high ecological validity, on the assumption that these stimuli are encountered on a daily basis. However, these stimuli were difficult to control due to their variability in perceptual and conceptual characteristics and individual differences in frequency of usage.

In addition to the difficulty mismatch already mentioned between the two dimensions – physical comparisons are easier and can be solved perceptually, whereas coin value comparisons require conceptual processing (see Chapter 2), other perceptual cues could have aided physical comparisons. Although the coin images had been modified to monochrome to prevent the use of colour cues during the comparison tasks, the amount of details on the coins varied, in particular, the £2 coin has a very complex design which could give subjects some advantage and they would be more likely to remember this coin then other ones on subsequent trials, hence judgement might have been made upon memory of previous trials rather than upon comparison per se.

Another perceptual characteristic which could not be controlled for was the shape of the coins. All the British coins are round, except the 20p and the 50p coins which are heptagonal, and this perceptual difference could also have affected subjects' judgements.

Furthermore, it was impossible to control for the subjects' usage of coins in everyday life. The increasing usage of credit cards, internet bank transfers, and other methods of electronic payment could mean differential experience in subjects with different preference in money handling. In addition, the later introduction of the £2 coin meant that subjects would have less experience with than other coins, and this could also have affected subjects' performance.

Nonetheless, the present experiment has provided a new way to investigate numerical information processing and shed new light onto the complicated effects of familiarity (or practice) on the Stroop phenomenon.

5 Parity Judgement Tasks

5.1 Introduction

Parity, a concept closely related to numerical magnitude, is another important property of numbers. The parity status of a number refers to whether it is odd or even. By definition, an even number is one that can be divided by two without any remainder, whereas an odd number is one that, when divided by two, leaves a remainder of one. Research in parity, in comparison to numerical magnitude, has not been as extensive. The present chapter considers the relationship between these concepts.

5.1.1 The "Odd" Effect and the Markedness Association of Response Codes (MARC) Effect

The current section considers the key findings of previous research relating to parity processing.

Hines (1990), using a simple parity judgement task, reported an "odd" effect – odd numbers were responded to slower than the even ones. He attributed this effect to the linguistic markedness of the adjective "odd" (Zimmer, 1964). However, Dehaene et al. (1993) failed to replicate the effect (Dehaene et al., 1993, Experiment 1). The difference could be explained by a closer look at the stimuli, the notations, and the measurements used in these experiments.

Hines (1990) reported an odd effect in terms of reaction time for number words (Hines, 1990, Experiment 5), but not for Arabic digits (Hines, 1990, Experiment 2). However, the latter did show an odd effect in terms of error rate. Dehaene et al.'s (1993, Experiment 1) failure to find an odd effect in terms of reaction time with Arabic digits (ranging from 0 to 9) therefore replicated Hines' (1990) finding. Intriguingly, the same authors reported an odd effect with both Arabic digits and French written verbal numerals in another experiment with numbers again ranging from 0 to 9 (Dehaene et al., 1993, Experiment 9).

Nuerk, Iversen, and Willmes (2004) explained the above inconsistent results in terms of the inclusion or exclusion of zero. While Hines (1990) employed stimuli from 2 to 9, Dehaene et al. (1993) used stimuli from 0 to 9. The 0 in Dehaene et al.'s (1993) Experiment 1 was responded to significantly slower than other even numbers (2, 4 and 8). This could have diluted any potential odd effect. Nuerk et al.'s (2004) explanation appears plausible but is inconsistent with Dehaene et al.'s (1993, Experiment 9) finding that 0 was responded to significantly slower than other even numbers 2, 4, 6, 8 while the odd effect was still present. Nonetheless, most researchers agree that zero is not a typical even number and should not be investigated along with other numbers (see Brysbaert, 1995; Fias, 2001; see also Armstrong, Gleitman & Gleitman, 1983, who showed that not all numbers are subjectively equally odd or even).

Willmes and Iversen (1995) reported that subjects responded faster in the even-right (odd-left) condition than in the reverse parity-response key configuration, i.e., even-left (odd-right). The authors considered "even" and "right" to be the antonyms of the marked adjectives "odd" and "left", and thus interpreted the observed Markedness Association of Response Codes (MARC) effect as a linguistic markedness congruency effect. The effect was stronger for written numerals than for Arabic digits, reflecting a stronger access to verbal-linguistic concepts such as markedness via verbal stimuli (see also Hines, 1990).

In Hines' (1990) Experiment 1, subjects had to decide whether two numbers had the "same" (non-marked) or "different" (marked) parity. The odd effect was very large (150-200 ms for young participants and about 300-400 ms for older participants) compared to that in the simple parity judgement task above (approximately 20 ms). Nuerk et al. (2004) attributed this huge difference to the importance of the congruency of the markedness attributes of stimulus and response in addition to the markedness of the stimulus per se; the markedness-incongruent odd-same condition was much slower than the markedness-congruent even-same condition.

5.1.2 The Availability of Parity Information

Sudevan and Taylor (1987) investigated the availability of parity information. On randomly intermixed trials, the same target digits had to be classified either as "odd versus even" or as "smaller than 6 versus larger than 5". The larger-smaller comparisons were consistently faster than parity judgements (see also Boles, 1986). Furthermore, during parity judgements, the larger-smaller status of the target digit interfered with the odd-even classification, suggesting an irrepressible activation of numerical comparison. These findings suggest that numerical magnitude is more readily available than parity.

The difference in availability between numerical magnitude and parity may be related to the developmental sequence of these concepts. Miller and Gelman (1983) studied the development using a number-similarity judgement task, in which they presented subjects with triads of single digits and asked them to judge which two were most and which two were least closely related to each other. They observed that parity had no influence until the sixth grade, whereas even kindergartners showed an early structuring of their mental number representations by numerical magnitude – kindergartners judged 2 to be more similar to 3 than to 7, but unlike adults, they did not find 2 more similar to 6 than to 5. This study thus highlights the developmental precedence of magnitude over parity.

Neuropsychological evidence also suggests that parity information may be less available than magnitude information. Dehaene and Cohen (1991) described a severely aphasic and acalculic patient, NAU, who could not read, memorize, or calculate with numbers in any exact way but remained able to perform numerical comparison and approximation. However, given his intact "analogue" magnitude representation, NAU was at chance level in parity judgement (55.3% errors with odd numbers and 32.4% errors with even numbers), suggesting that the parity judgement is a more demanding "exact" task compared to numerical comparison which is seen here as an approximation task based on the use of an analogue representation of numbers. The authors explained the findings in terms of a dissociation between impaired exact processing (e.g., parity judgement) versus preservation of approximation abilities (e.g., numerical comparison). This view relies on two mechanisms, where exact processing manipulates symbolic representations of numbers but approximate processing requires numbers to be converted to analogue magnitude representations. However, according to Zorzi and Butterworth's (1999) view on discrete numerosity representations, numerical comparison would be considered an exact process which compares two discrete numerosity representations.

On the other hand, one could also consider McCloskey's single-route model consisting of just a common store of internal representations of numbers. If parity information is normally accessed via magnitude representation (which, according to Zorzi & Butterworth, 1999, is discrete rather than analogue), then NAU's poor parity judgement can be interpreted as a deficit in the pathway between numerical magnitude and parity (see a, Figure 5.1). As an alternative possibility, parity information may be retrieved directly from memory, thus NAU's performance can be interpreted as impaired memory of parity information (see b, Figure 5.1) or impaired pathway to the memory of parity information (see c, Figure 5.1).

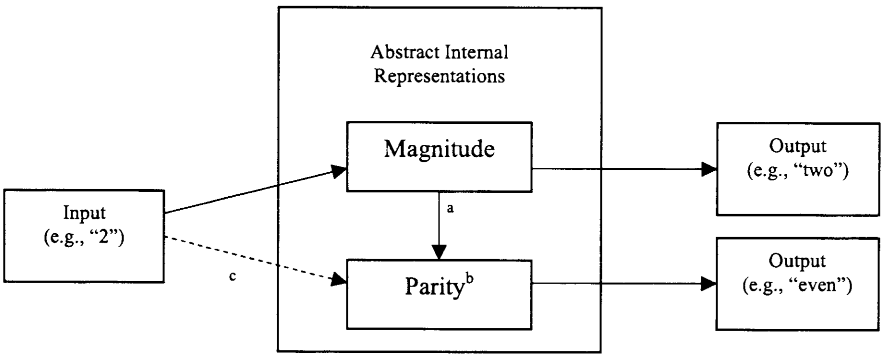


Figure 5.1 Proposed conceptualisation of the internal representations of and the pathways between numerical magnitude and parity, and the possible deficits in patient NAU; the dotted line indicates a weaker pathway based on the suggestion that "numerical magnitude is more readily available than parity information" (Dehaene et al., 1993)

Göbel et al. (2001b) reported impaired performance when the left angular gyrus was under rTMS in a number comparison task, where subjects had to decide whether a two-digit number was larger or smaller than 65, suggesting that this region was involved in processing numerical magnitudes. This finding has been replicated by Rusconi, Walsh, and Butterworth (2005, Experiment 2), in a numerical magnitude judgement task, where subjects had to decide whether or not two numbers presented were both larger than 5. Furthermore, Rusconi et al. (2005) reported no disruption in a parity judgement task, where subjects had to judge whether or not two numbers presented had the same parity, under the same rTMS condition. These findings implicate the left angular gyrus' involvement in numerical magnitude processing and suggest that the parity judgement task (Rusconi et al., 2005) might be solved without deliberate access to numerical magnitude. However, it must be noted that numerical magnitude representation per se is assumed to be localized in the bilateral intra-parietal sulci (Pinel et al., 2004).

The above findings, however, should not be interpreted to imply that numerical magnitudes have no influence on parity judgements; in fact, evidence has suggested the opposite (Dehaene et al., 1993; see also Andres, Davare, Pesenti, Oliver, & Seron, 2004). As shown in Experiments 2 and 3b of the present thesis, numerical magnitudes were autonomously processed when two numbers were presented under task-irrelevant conditions; this autonomous process might have occurred and influenced parity judgements in Rusconi et al's (2005) experiment where two numbers were presented. Experiments 5a and 5b were designed to explore this possibility.

An important point to note is that in Rusconi et al. (2005, Experiment 2), the mean reaction time for numerical magnitude judgement task was significantly slower than that for the parity judgement task. This contrasts with Sudevan and Taylor's (1987) finding where the mean reaction time pattern was in the opposite the direction. These findings indicate that reaction time patterns varied across the experimental paradigms even when the same dimensions (numerical magnitude and parity) were employed. Reaction time therefore could not be used reliably to make inferences on the availability of numerical magnitude and parity information.

5.1.3 Theoretical Accounts of Parity Information Processing

Early models of number processing have often neglected the processing of parity information; not until the 1990's had research started to address this aspect specifically (e.g., Clark & Campbell, 1991; Dehaene et al., 1993).

According to Clark and Campbell (1991), a mental division by 2 is performed during parity judgement, "odd and even are in fact defined and presumably determined by numerical calculations (e.g., multiple or non-multiple or 2)." In this view, parity judgement relies on simple procedures of mental calculation. Such an account predicts a problem size effect (e.g., Clapp, 1924), i.e., an increase in reaction time with increasing numerical magnitude of the target number during parity judgements.

Although the parity status of a number is mathematically defined in terms of its divisibility by 2, Dehaene et al. (1993) argued that it does not necessarily imply that the concept of parity is mentally represented as such. These authors proposed alternative strategies to determine the parity status of numbers, for example, "Subjects might use the rule that the numbers ending in 0, 2, 4, 6, or 8 are even, and all others are odd."

In two experiments (Dehaene et al., 1993, Experiments 1 & 2) which required subjects to judge the parity of one- and two-digit numbers respectively, Dehaene et al. (1993) examined the feasibility of the two strategies in determining the parity status of a number – whether it is via mental calculations or retrieval from memory. The authors reported that "parity judgment times did not generally increase as a function of the target number" (Dehaene et al., 1993, Experiment 1), contrary to what Clark and Campbell's (1991) account would predict. However, it should be noted that numerical magnitudes did have a certain degree of influence on parity judgements in Dehaene et al.'s experiment (1993, Experiment 1).

Dehaene et al. (1993, Experiment 1) reported some intriguing findings separately for odd and even numbers. Within the even category, 2, 4, and 8 were classified faster than the other even numbers 0 and 6, and the authors attributed this difference to the salient mental category – the powers of 2, consistent with Shepard, Kilpatrick, and Cunningham's (1975) finding. Within the odd category, 3, 5, and 7 were classified faster than 1 and 9, and the authors explained the difference in terms of prime and non-prime numbers, "The numbers 3, 5, and 7 are prime numbers, whereas 1 and 9 are not. Prime numbers may evoke a fast odd response, because only one prime number (2) is even." These findings clearly indicate that numerical magnitudes had a certain degree of influence on parity judgements. However, due to an absence of an increase in reaction time with numerical magnitude, Dehaene et al. (1993) interpreted these findings as support for the memory retrieval account of parity information.

Moreover, Berch, Foley, Hill, and McDonough Ryan (1999) had children judge the parity of single digits and reported that from fourth grade (mean age of 9.8 years) on, parity information was retrieved from memory. At third grade (mean age of 9.2 years), however, children appeared to adopt a skip-counting strategy (at least for even numbers). The authors concluded that school-age children did not employ a mental calculation strategy to determine parity status.

Experiments 5a and 5b of the present thesis made use of Hines' (1990, Experiment 1) same/ different parity judgement paradigm, incorporating into it various levels of numerical distance to investigate numerical magnitude processing in addition to parity information processing. The task required subjects to judge whether two numbers (Arabic digits and English written verbal numerals in the two experiments respectively) were of the same parity or not. Previous research findings have suggested that numerical magnitude is more available than parity information. It was therefore predicted that during parity judgements of number pairs, reaction times would vary according to the numerical distance between the numbers.

Results of Experiments 2 and 3b demonstrated refined autonomous processing of numerical magnitudes under task-irrelevant conditions. When numerical magnitude was manipulated as the task-irrelevant dimension, a reversed distance effect (indicated by a positive linear trend) was observed. In other words, a pair of numbers with a larger numerical distance between them interfered more strongly with the relevant (physical comparison) task than a pair of numbers with a smaller numerical distance. Applying this to the current parity judgement task, the autonomous comparison of the task-irrelevant numerical magnitudes could impede the relevant parity judgement task, making it difficult to perform, and an increase in reaction time during parity judgements was predicted as the task-irrelevant numerical distance between the numbers increased, i.e., a positive linear trend.

A second hypothesis stems from Dehaene et al.'s (1993) memory retrieval account. It assumes that parity information is received from memory, therefore predicts faster reaction times with numbers which are more familiar, e.g., even numbers (Dehaene et al. 1993). These authors pointed out that "the series 2, 4, 6, 8,... is better known than the series 1, 3, 5, 7,... because it is practiced more often, for instance, when counting by twos. Indeed, recitation of the even series is sometimes preserved in brain-lesioned aphasic patients who fail to recite the odd series" (e.g., Dehaene & Cohen, 1991). Experiments 5a and 5b required subjects to judge whether or not two numbers presented had the same parity, and the memory retrieval account would predict faster reaction times with trials containing two even numbers than those containing one odd and one even numbers, and those containing two odd numbers.

Experiments 5a and 5b also tested a prediction arising from the SNARC effect which refers to the preferential rightward response with large numbers and leftward response with small numbers (Dehaene et al., 1993; see also Fias et al., 1996). This effect has been interpreted to reflect a left-to-right mental organisation of numbers. In half of the trials in Experiments 5a and 5b, the smaller number of a pair would appear on the left, but in the other half, the smaller one would appear on the right. A left-to-right mental organisation of numbers would be indicated by faster mean reaction time with trials containing a small number on the left than on the right, thus a stimulus-side effect.

5.2 Experiment 5a: Parity Judgement Task with Arabic Digits

5.2.1 Methods

5.2.1.1 Task

The task required subjects to judge, in each trial, whether a pair of numbers (Arabic digits) had the same or different parity via a button key press. Their reaction times and responses were recorded.

The task was computer-based. A program written in Cogent (running on a MATLAB Version 6.1 platform) was used.

5.2.1.2 Stimuli

In each trial, two digits appeared simultaneously in white on a black background. Each presentation lasted 100 ms, and there was an interval of 3000 ms before the subsequent trial.

The stimuli were Arabic digits (1 to 9 excluding 5, so that there were equal number of odd and even numbers) in font Arial. The pairs used were: 1 2, 2 3, 8 9 (for numerical distance of 1); 1 3, 4 6, 7 9 (for numerical distance of 2); 1 4, 4 7, 6 9 (for numerical distance of 3); 2 6, 3 7, 4 8 (for numerical distance of 4); 1 6, 2 7, 3 8 (for numerical distance of 5); 1 7, 2 8, 3 9 (for numerical distance of 6). Each of these pairs was presented twice – once with the smaller number of the left, and once with the smaller number on the right. The total number of trials in each task was 36, half of which had the correct response as "same", and the rest were "different". Half the subjects used their left index finger (on the "F" key of a qwerty keyboard) to make the "same" response and their right index finger (on the "J" key of a qwerty keyboard) to make the "different" response, and the opposite applied to the other half of the subjects.

It is important to note that when the difference between the numbers is 1, 3, or 5, the correct response would always be "different", whereas when the difference is 2, 4, or 6, the correct response would always be "same".

Stimuli were presented in a pseudorandom order with the following constraints to avoid carryover effects: (1) the same digit did not appear in consecutive trials, (2) the correct response did not appear on the same side (left or right) for more than three consecutive trials, (3) the numerical distance was not the same for more than two consecutive trials, and (4) the smaller number did not appear on the same side for more than four consecutive trials.

5.2.1.3 Subjects

There were 32 subjects (20 females and 12 males), age ranged 21 to 38 (mean = 25.8 years, standard deviation = 4.7 years). All subjects had normal or corrected-to-normal eyesight.

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 3000 ms after the stimulus offset. One left-handed subject and two subjects with error rates over 20% were excluded from further analyses. The mean error rate (N = 29) was 8.53%.

5.2.2 Results

Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were removed.

ANOVAs were used to analyse mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

When mean reaction times were plotted against numerical distance between the numbers, a zigzag line in a general upward direction emerged (see Figure 5.2). There appears a difference between trials with "different" response (i.e., when the distance was 1, 3, or 5) and those with "same" response (i.e., when the distance was 2, 4, or 6) – the former appear to have longer mean reaction times than the latter. Hence, the "different" and "same" trials were separated for further analyses.

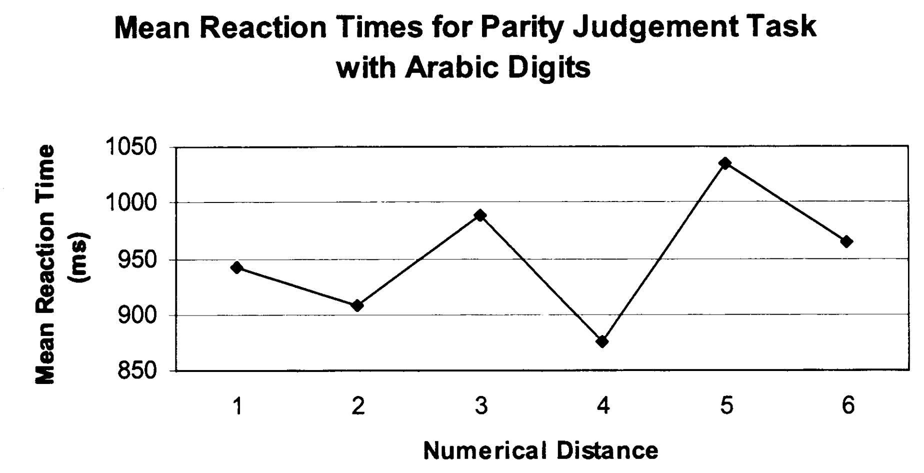


Figure 5.2 Mean reaction times (ms) for parity judgement task with Arabic digits (Experiment 5a)

A 3 x 2 x 2 repeated-measures ANOVA was used, where the mean reaction times were decomposed into different factors: distance (the numerical distance between the numbers was either small – 1 and 2, medium – 3 and 4, or large – 5 and 6), parity (the numbers had either same or different parity), and stimulus side (the smaller number appeared either on the left or on the right). The ANOVA revealed a significant main effect of distance ($F_{(2, 56)} = 4.60$, $p < 0.020$), a significant main effect of parity ($F_{(1, 28)} = 23.26$, $p < 0.001$) – that "same" responses were significantly faster than "different" responses (929 ms vs. 1007 ms respectively), and a non-significant main effect of stimulus side ($F_{(1, 28)} < 1$, *n.s.*). All interactions were non-significant. Tests of within-subjects contrasts revealed a significant linear trend ($F_{(1, 28)} = 5.92$, $p < 0.050$) and a non-significant quadratic trend ($F_{(1, 28)} = 2.45$, *n.s.*) for the factor distance. See Figure 5.3 for a graphical representation of the mean reaction time patterns at each level of parity.

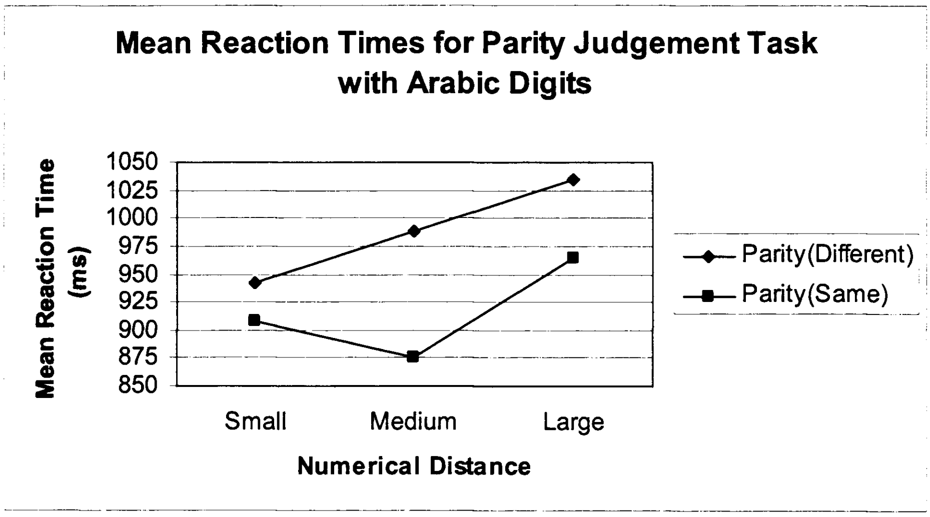


Figure 5.3 Mean reaction times (ms) for "same" and "different" trials with Arabic digits (Experiment 5a)

Paired-samples t-tests were conducted to further investigate the difference in mean reaction time between "same" and "different" responses. They revealed a non-significant difference between trials with two odd numbers and trials with one odd and one even number ($t_{(28)} = -1.06$, *n.s.*), but a significant difference between trials with one odd and one even number and trials with two even numbers ($t_{(28)} = 6.78$, $p < 0.001$); mean reaction times were 1036 ms, 1008 ms, and 840 ms respectively (see Figure 5.4).

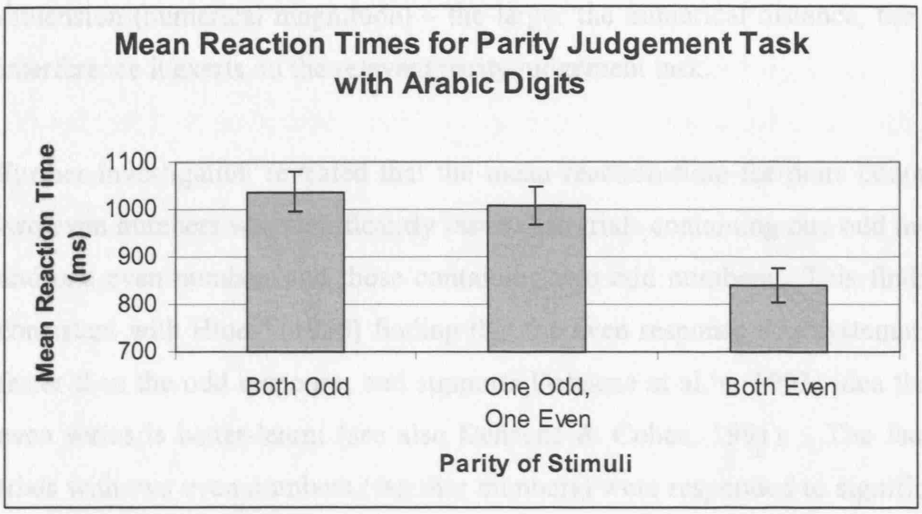


Figure 5.4 Mean reaction times (ms) for "both odd," "one odd, one even," and "both even" trials with Arabic digits (Experiment 5a)

5.2.3 Discussion

The present experiment set out to investigate numerical magnitude processing during parity judgements. Subjects were required to judge whether a pair of numbers (Arabic digits) presented had the same parity status or not. By varying the numerical distances between the number pairs, it was possible to examine whether or not the task-irrelevant dimension – numerical magnitude – was autonomously processed. The paradigm also allowed the investigation into the strategies used in processing parity information.

During parity judgements, the mean error rate was found to be low (< 9%). Mean reaction times for parity judgements with Arabic digits varied with numerical distance between the numbers.

When mean reaction times were plotted against numerical distance between the numbers, a zigzag line in a general upward direction (Figure 5.2) emerged. Further reaction time analyses revealed a significantly faster mean judgement time when two numbers had the same than when they had different parity, and that both the "same" and the "different" trials showed a significant positive linear trend with increasing numerical distance between the two numbers. The positive trends can be interpreted in terms of interference from the task-irrelevant dimension (numerical magnitude) – the larger the numerical distance, the more interference it exerts on the relevant parity judgement task.

Further investigation revealed that the mean reaction time for pairs containing two even numbers was significantly faster than trials containing one odd number and one even number, and those containing two odd numbers. This finding is consistent with Hines' (1990) finding that the even response was systematically faster than the odd response, and supports Dehaene et al.'s (1993) idea that the even series is better-learnt (see also Dehaene & Cohen, 1991). The fact that trials with two even numbers (familiar numbers) were responded to significantly faster than other trials supports the memory retrieval account proposed by Dehaene et al. (1993).

The finding that trials containing two even numbers were responded to significantly faster than trials containing two odd numbers could also be interpreted as a linguistic markedness congruency effect. Both trial types required the correct response "same" which is non-marked; trials with two even (non-marked) numbers were therefore linguistically congruent, whereas those with two odd (marked) numbers were linguistically incongruent. The longer mean reaction time with the latter could thus be taken to reflect the informational conflict at a linguistic level.

Statistically, mean reaction times showed a significant linear trend (without a significant quadratic trend) with both "same" and "different" trials as numerical distance increased. However, the plot of the "same" trend was not strictly linear – mean reaction for pairs with a numerical distance of 4 was faster than that for pairs with a numerical distance of 2. This may be explained in terms of the distribution of number pairs. Out of the three number pairs used for each level of numerical distance, those used for numerical distance of 4 (i.e., 2 6, 3 7, 4 8) contained more even numbers (4) than those used for any other level of distance (2 for numerical distances 2 and 6). Given the larger proportion of even numbers, the faster mean reaction time for pairs with a numerical distance of 4 is consistent with Hines' (1990) finding and with Dehaene et al.'s (1993) suggestion that even numbers are better-learnt than odd numbers.

The present paradigm is rather difficult compared to Dehaene et al.'s (1993) Experiment 1 where subjects only had to determine the parity of a single digit presented on its own, so subjects in the current experiment might have employed multiple strategies. There is already evidence supporting memory retrieval in accessing parity information – two even numbers were responded to significantly faster than other trial types. However, this memorial strategy alone cannot explain all the data. Other strategies might also have been in operation, for example, one might use the rule of thumb that when numbers are adjacent to each other in the numerical sequence (i.e., when the numerical distance is 1), they must have different parity statuses.

In summary, current findings highlight the complexity of parity information processing and suggest that multiple strategies might be in place during complex parity judgements.

Findings of the current experiment provide evidence for refined numerical information processing when two numbers were presented. In this parity judgement task, not only were subjects faster in responding to "same" than to "different" trials, they also showed a reversed distance effect with both of these trial types. Since numerical magnitude was manipulated as the task-irrelevant dimension, the reversed distance effects observed reflect autonomous processing. However, there is still uncertainty as to whether parity information can be extracted without the number semantics (numerical magnitudes) since only a neuropsychological dissociation has been reported in the direction of preserved numerical magnitudes and impaired parity information (Dehaene & Cohen, 1991) but not vice versa, one can only draw the conclusion that numerical magnitude is more available than parity information.

No evidence was found in the current experiment for the stimulus-side effect – a prediction based upon the SNARC effect (Dehaene et al., 1993). If one adopts the view that the SNARC effect reflects semantic (in this case, magnitude) access (Fias, 2001), then the absence of a stimulus-side effect may be taken to indicate a lack of access to numerical magnitudes. However, the reversed distance effects observed in the present experiment have provided evidence otherwise, rendering the stimulus-side effect an unreliable indicator of semantic (magnitude) processing. On the other hand, if a stimulus-side effect had been observed, it could be interpreted to reflect ordinal rather than semantic (magnitiude) representations of numbers (see Gevers et al., 2003). In summary, the absence of a stimulus-side effect does not support a left-to-right mental organisation of numbers.

5.3 Experiment 5b: Parity Judgement Task with Written Verbal Numerals

5.3.1 Methods

5.3.1.1 Task

The task required subjects to judge, in each trial, whether a pair of numbers (English written verbal numerals) had the same or different parity via a button key press. Their reaction times and responses were recorded.

The task was computer-based. A program written in Cogent (running on a MATLAB Version 6.1 platform) was used.

5.3.1.2 Stimuli

The only difference between the current experiment and Experiment 5a was that stimuli were English written verbal numerals instead of Arabic digits. All the criteria for stimulus selection were exactly the same as those in Experiment 5a.

5.3.1.3 Subjects

There were 12 subjects (9 females and 3 males), age ranged 19 to 34 (mean = 25.0 years, standard deviation = 4.9 years). All subjects were native English speakers and had normal or corrected-to-normal eyesight.

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 3000 ms after the stimulus offset. One subject with error rates over 20% was excluded from further analyses. The mean error rate (N = 11) was 6.57%.

5.3.2 Results

Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were removed.

ANOVAs were used to analyse mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

When mean reaction times were plotted against numerical distance between the numbers, a zigzag line in a general upward direction emerged (see Figure 5.5). There appears to be a difference between trials with an odd response (i.e., when the distance was 1, 3, or 5) and those with an even response (i.e., when the distance was 2 or 4, but not 6), – the former appear to have longer mean reaction times than the latter. Hence, the "different" and "same" trials were separated for further analyses.

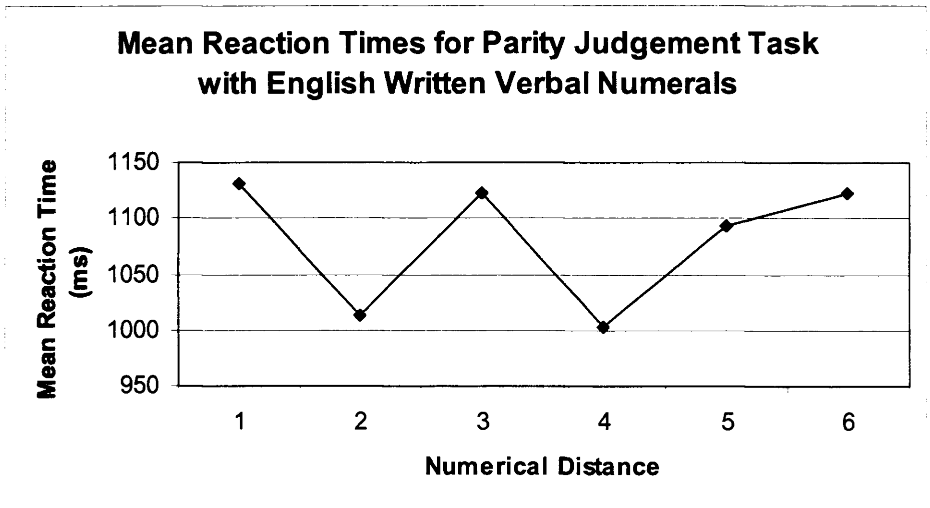


Figure 5.5 Mean reaction times (ms) for parity judgement task with English numerals (Experiment 5b)

A 3 x 2 x 2 repeated-measures ANOVA was used, where the mean reaction times were decomposed into different factors: distance (the numerical distance between the numbers was either small – 1 and 2, medium – 3 and 4, or large – 5 and 6), parity (the numbers had either same or different parity), and stimulus side (the smaller number appeared either on the left or on the right). The ANOVA revealed no significant main effect (all *n.s.*). The only significant interaction was distance x parity ($F_{(2, 20)} = 3.83$, $p < 0.050$). All other interactions were non-significant.

Further analyses were conducted on mean reaction times at each level of parity. When the parity was different, test of within-subjects contrasts revealed a non-significant linear ($F_{(1, 10)} = 1.31$, *n.s.*) and a non-significant quadratic ($F_{(1, 10)} < 1$, *n.s.*) trends. When the parity was the same, test of within-subjects contrasts revealed a significant linear trend ($F_{(1, 10)} = 16.21$, $p < 0.005$) and a non-significant quadratic trend ($F_{(1, 10)} = 1.71$, *n.s.*). See Figure 5.6 for a graphical representation of the mean reaction time patterns at each level of parity.

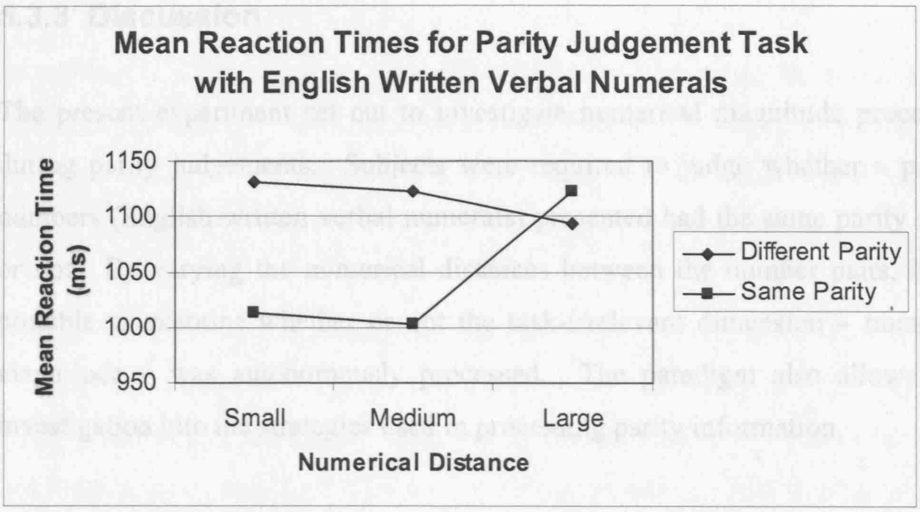


Figure 5.6 Mean reaction times (ms) for "same" and "different" trials with English numerals (Experiment 5b)

Paired-samples t-tests were conducted to further investigate the difference in mean reaction time between "same" and "different" responses. They revealed a non-significant difference between trials with two odd numbers and trials with one odd and one even number ($t_{(10)} = -0.64$, *n.s.*), but a significant difference between trials with one odd and one even number and trials with two even numbers ($t_{(10)} = -3.94$, $p < 0.005$); mean reaction times were 1176 ms, 1151 ms, and 990 ms respectively (see Figure 5.7).

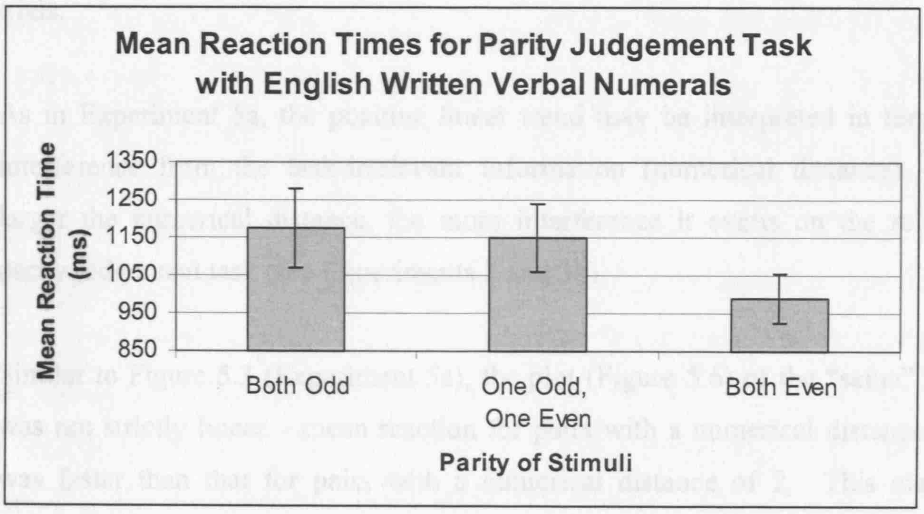


Figure 5.7 Mean reaction times (ms) for "both odd," "one odd, one even," and "both even" trials with English Numerals (Experiment 5b)

5.3.3 Discussion

The present experiment set out to investigate numerical magnitude processing during parity judgements. Subjects were required to judge whether a pair of numbers (English written verbal numerals) presented had the same parity status or not. By varying the numerical distances between the number pairs, it was possible to examine whether or not the task-irrelevant dimension – numerical magnitude – was autonomously processed. The paradigm also allowed the investigation into the strategies used in processing parity information.

Similar to Experiment 5a, the mean error rate for parity judgements was low (< 7%). Mean reaction times for parity judgements with English written verbal numerals varied with numerical distance between the numbers.

When mean reaction times were plotted against numerical distance between the numbers, a zigzag line in a general upward direction emerged (see Figure 5.5). Unlike Experiment 5a, mean reaction times for "same" and "different" trials were not significantly different from one another. However, the factor parity interacted significantly with numerical distance. Further analyses revealed a significant positive linear trend – an increase in reaction time with increasing numerical distance – was observed with the "same" trials, but not the "different" trials.

As in Experiment 5a, the positive linear trend may be interpreted in terms of interference from the task-irrelevant information (numerical distance) – the larger the numerical distance, the more interference it exerts on the relevant parity judgement task (see Experiments 2 and 3b).

Similar to Figure 5.3 (Experiment 5a), the plot (Figure 5.6) of the "same" trend was not strictly linear – mean reaction for pairs with a numerical distance of 4 was faster than that for pairs with a numerical distance of 2. This may be explained in terms of the distribution of number pairs. Out of the three number pairs used for each level of numerical distance, those used for numerical distance of 4 (i.e., 2 6, 3 7, 4 8) contained more even numbers (4) than those used for any other level of distance (2 for numerical distances 2 and 6). Given the larger proportion of even numbers, the faster mean reaction time for pairs with a numerical distance of 4 is consistent with Hines' (1990) finding and with Dehaene et al.'s (1993) suggestion that even numbers are better-learnt than odd numbers.

Further investigation, comparing different trial types, revealed the same reaction time pattern as observed in Experiment 5a – those containing two even numbers was responded to significantly faster than those containing one odd number and one even number, and those containing two odd numbers. This finding supports Dehaene et al.'s (1993) idea that the even series is better-learnt (see also Dehaene & Cohen, 1991).

As discussed in Section 5.2.3, the finding that trials containing two even numbers were responded to significantly faster than trials containing two odd numbers could also be interpreted as a linguistic markedness congruency effect. Moreover, the size of this effect observed in the current experiment was comparable to that observed in Experiment 5a (differences in mean reaction time were 186 ms and 196 ms respectively). This, however, is inconsistent with Willmes and Iversen's (1995) finding that the effect was stronger for written numerals than for Arabic digits (see also Nuerk et al., 2004), and therefore does not provide support for a stronger access to verbal-linguistic concepts such as markedness via verbal stimuli.

As in Experiment 5a, no evidence was found for the stimulus-side effect – a prediction based upon the SNARC effect (Dehaene et al., 1993). In other words, trials containing a small number on the left were not responded to significantly faster than those containing a smaller number on the right. Once again, this finding (1) reinforces the suggestion that the stimulus-side effect cannot be used as a reliable indicator of semantic (magnitude) processing, and (2) does not support a left-to-right mental organisation of numbers.

The current experiment, together with Experiment 5a, highlights the complexity of parity information processing and suggests that multiple strategies might be in place during complex parity judgements. There is some (limited) evidence to suggest that numerical magnitudes were processed autonomously during parity judgements with English written verbal numerals.

6 Processing Numerical Magnitude and Numerosity Information

6.1 Introduction

The comparison tasks discussed in Chapters 2, 3, and 4 are of a very different nature: physical size comparison tasks can be solved perceptually, whereas numerical magnitude and coin value comparison tasks require conceptual processing. Given the difference in nature, these dimensions appear not very well-matched. In another numerical Stroop variant, which involves numerosity naming and reading Arabic numerals, both dimensions require conceptual processing.

Windes (1968) first documented the interference effect in numerosity naming and reading Arabic digits. When subjects were asked to enumerate the number of items being displayed (i.e., the numerosity), slower reaction time was observed when the numerical values represented by the digits were incompatible with numerosity. For example, it would take longer to respond to the incongruent trial (2 2 2) compared to the congruent trial (2 2). Neutral trials, where the items have no numerical value (e.g., X X X), could also be included in this type of paradigm. Windes' (1968) finding was replicated by several studies (e.g., Flowers et al., 1979; Shor, 1971). Using a card-sorting task in which participants were required to order a group of cards according to the number of symbols printed on them, Morton (1969) observed interference with both Arabic digits and written verbal numerals.

In a thorough study by Pavese and Umiltà (1998), the numerical distance effect was investigated using this type of enumeration Stroop paradigm. With a strategic design, they examined the numerical distance effect in both small and large numerosity conditions. In the "small" condition, 5 digits (1 to 5) and 5 numerosities (1 to 5) were used, whereas in the "large" condition, the digits used were 5 to 9 and numerosities were also 5 to 9. Pavese and Umiltà (1998) observed that enumeration of small numerosities required significantly less time than that of larger ones. In the "small" condition, reaction time was observed to increase minimally but significantly with numerosity. However, in the "large" condition, reaction time was found to increase sharply and significantly with numerosity. These findings support the idea that there are two enumeration mechanisms: subitizing which involves simple pattern-matching based on recognition and counting which relies on the existence of counting knowledge, represented in the form of an ordered sequence of number facts (see also Kaufman et al., 1949; Mandler & Shebo, 1982; Peterson & Simon, 2000).

Peterson and Simon (2000) argued that for small numerosities, where the number of possible patterns is sufficiently small, all patterns are referenced frequently enough that they become highly activated in memory, and thus reaction times are short for this type of enumeration. On the other hand, the number of possible patterns is much greater for larger numerosities, and therefore, each individual pattern is seen much less frequently, hence the reliance on a more algorithm-based mechanism, i.e., counting, giving rise to longer reaction times. It is important to emphasise that empirical data for enumeration of numerosities in the counting range do not suggest a strict usage of an item-by-item counting procedure – although counting is predominantly algorithm-based, subpattern recognition (i.e., pattern recognition of smaller chunks within the subitizing range) is also involved (Frick, 1987; Peterson & Simon, 2000; Piazza, Mechelli, Butterworth, & Price, 2002; Shrager, Klahr, & Chase, 1982).

With regard to the numerical distance effect, Pavese and Umiltà (1998) found that reaction times for incongruent close trials were significantly slower than those for incongruent distant trials (e.g., 3 3 3 3 and 1 1 1 1 respectively) in both conditions with small and large numerosities. In the experiment, they employed a design where the stimulus array was displayed until a response was given. Such a procedure might have allowed participants to process task-irrelevant digit identity at a later stage, i.e., after the enumeration process has been completed. So, Pavese and Umiltà (1999) carried out another experiment, employing a slightly different design so that the stimulus array was briefly presented (200 ms) and masked. Even with such a short presentation time, the numerical distance effect was replicated – incongruent close digits were enumerated more slowly than incongruent distant digits. Such a finding suggests that the numerical distance effect depends on a rapid and autonomous (rather than late) activation of numerical magnitude information, even when the information is task-irrelevant.

The numerical distance effect observed in Pavese and Umiltà's (1998, 1999) experiments challenges Tzelgov et al.'s (1992) theoretical account based on Logan's (1988) two-process theory of skilled performance. According to Tzelgov et al. (1992), numerical magnitude comparisons are dominated by a slow intentional process, and when task-irrelevant, "only a crude, dichotomous representation of numerical size was encoded". This account would predict a significant Stroop effect without a numerical distance effect since the latter reflects refined information processing of numerical magnitudes. Thus, the numerical distance effect observed in Pavese and Umiltà's (1998, 1999) enumeration Stroop tasks can be seen as inconsistent with Tzelgov et al.'s (1992) account. However, it must be noted that the numerical distance effect reported by Pavese and Umiltà (1998, 1999) was a mere difference in mean reaction time between numerically distant and close pairs, rather than a significant linear decrease with increasing numerical distance. Consequently, caution must be taken in interpreting the reported effect to reflect refined numerical magnitude processing.

6.2 Methodological Issues Concerning the Enumeration Paradigm

Although the enumeration paradigm has the advantage of employing two dimensions of similar nature (namely, numerical magnitude and numerosity) both of which require processing on a cognitive level, there are major drawbacks concerning the paradigm.

6.2.1 Response Preparation and Interference Process Are Confounded

Firstly, the enumeration paradigm (Pavese & Umiltà, 1998) suffers the same criticism by Zysset et al., (2001) to the traditional colour-word Stroop paradigm (Stroop, 1935), "generating the verbal response to match a stimulus is interfered by the second dimension of the stimulus (or the dimension of a second stimulus). Response preparation and the interference process itself are confounded by this." In the traditional colour-word Stroop paradigm, for example, generating the verbal response "red" to match one dimension, namely colour, to the stimulus "GREEN" is interfered with by the second dimension of the stimulus, namely the identity of the word. Likewise in the enumeration Stroop paradigm, generating the verbal response "four" to match one dimension, namely numerosity, to the stimulus set "3 3 3 3" is interfered by the second dimension of the set, namely the identity of the digits. Even in experiments where responses were made via key pressing to the corresponding numerosity (e.g., Bush et al., 1998, 1999), the problem of responding to match only one of the two stimulus attributes was not eliminated.

To combat this problem, Zysset et al. (2001) incorporated a matching process in the traditional colour-word Stroop task. The modified task required subjects to compare, not just one, but two stimulus attributes; they had to make a judgement as to whether the colour of the upper word corresponded to the identity of the lower word (see sample stimuli in Table 6.1).

Table 6.1 Sample stimuli of Zysset et al.'s (2001) experiment

Correct Response	Congruent	Neutral	Incongruent
No	RED BLUE	XXXX BLUE	GREEN BLUE
Yes	BLUE BLUE	XXXX BLUE	GREEN BLUE

Table 6.2 shows a comparison between traditional colour-word Stroop task (Stroop, 1935) and Zysset et al.'s (2001) colour-word matching task. In the former, interference and response preparation are confounded since interference is between two different dimensions (colour and word), but response preparation is always on the same dimension (word), whereas the introduction of a matching process in Zysset et al.'s (2001) task disentangles the two processes, so that interference occurs within the same dimension (between words), but response competition (between manual responses) is separated.

Table 6.2 Comparisons between the traditional colour-word Stroop task (Stroop, 1935) and Zysset at al.'s (2001) colour-word matching task

	Traditional Colour-Word Stroop Task (Stroop, 1935)		**Zysset et al.'s (2001) Colour-Word Matching Task**	
Sample Stimuli	YELLOW	GREEN	YELLOW BLUE	GREEN BLUE
Relevant Dimension(s)	colour	colour	colour (upper row) and word (lower row)	colour (upper row) and word (lower row)
Irrelevant Dimension	word	word	word (upper row)	word (upper row)
Interference	"yellow" → colour red (i.e., word to colour)	"green" → colour red (i.e., word to colour)	"yellow" → "blue" (between words)	"green" → "blue" (between words)
Suppression of Irrelevant Information	"yellow"	"green"	"yellow"	"green"
Response Suppression	"yellow"	"green"	"match" key	"match" key
Correct Response Preparation	"red"	"red"	"mismatch" key	"mismatch" key
Response Competition	between "yellow" and "red"	between "green" and "red"	between "match" and "mismatch" keys	between "match" and "mismatch" keys

While disentangling response preparation from the interference process with the introduction of a matching process, Zysset et al. (2001) still managed to produce a Stroop interference effect, suggesting that the main source of interference in their colour-word matching task was the same as in the traditional colour-word Stroop task. According to the authors, "the matching process results in a manual response process, which is unaffected by the interfering dimensions of the stimuli." In this way, "response preparation is kept constant", so the lack of substantial activation in the anterior cingulate cortex (ACC) when comparing between conflict and non-conflict trials was not surprising. Instead, the authors argued that "other Stroop tasks used so far contrasted the response preparation process which is confounded with the interference process" and that "the ACC activation reported in previous studies reflects the degree of response conflict and not interference per se." They concluded that "regions along the IFS appear to be involved in solving interference effect and task management." (See Chapter 3 for a discussion on the brain regions involved in conflict resolution.)

6.2.2 Response Preparation and Response Suppression Are Not Controlled

Zysset et al.'s (2001) paradigm had a second advantage, over the traditional colour-word Stroop task (Stroop, 1935) and the enumeration Stroop task (Pavese & Umiltà, 1998), for keeping response suppression constant.

In a task where subjects make verbal responses, response preparation and response suppression cannot be well-controlled. Verbal response preparation is different for each different digit (or colour in the case of colour-word Stroop task), and more importantly, it is different for each suppressed digit (or colour, see Table 6.2). For example, stimulus sets, "2 2 2" and "5 5 5" both require a correct answer of "three", however, one needs to suppress the preparation for "two" in the former, but "five" in the latter. The introduction of a matching process which allows subjects to response by key presses removes this problem, since the suppressed response (the "mismatch" key) is separated from the suppressed information (see Table 6.2).

A new experimental design could be devised analogically from Zysset et al.'s (2001) paradigm (see Table 6.3).

Table 6.3 Sample stimuli of a proposed experimental design for the investigation of numerical magnitude and numerosity information processing

Correct Response	Congruent	Incongruent
No/ Mismatch	3 3 3 # # # # #	5 5 5 # # #
Yes/ Match	3 3 3 # # #	5 5 5 # # # # #

To maintain Zysset et al.'s (2001) key design aspect, the task would require subjects to compare two stimulus attributes, namely numerical magnitude (or the identity of the digits) and numerosity, i.e., the subjects have to indicate, via key pressing, whether the numerical magnitude (or identity) of the digits in the upper row matches the numerosity (the number of items) in the lower row. However, there would be no adequate neutral trials, where the stimuli have no numerical magnitude (e.g., X X X). Furthermore, the source of interference here would be the numerosity in the upper row, thus autonomous processing of numerical magnitude could not be examined.

An alternative task – the one used in Experiments 6a and 6b – required subjects to match the two sets of numerosity (see Table 6.4). In this way, however, interference and response preparation remain confounded, since subjects would only be comparing one stimulus attribute, namely numerosity, while trying to ignore information from the other stimulus attribute, namely the numerical magnitude (or identity) of the digits. Nevertheless, this newly devised paradigm has the advantage of controlling for response suppression (see Table 6.5).

Table 6.4 Sample stimuli (Experiments 6a and 6b)

Correct Response	Congruent	Neutral	Incongruent
No/ Mismatch	3 3 3 # # # #	X X X # # # #	5 5 5 # # # #
Yes/ Match	3 3 3 # # #	X X X # # #	5 5 5 # # #

Table 6.5 Comparisons between Pavese and Umiltà's (1998) enumeration Stroop task and the currently proposed numerosity-matching task (Experiments 6a and 6b)

	Pavese and Umiltà's (1998) Paradigm		Present Paradigm	
Sample Stimuli	5 5 5	2 2 2	5 5 5 # # # #	2 2 2 # # # #
Relevant Dimension	numerosity	numerosity	numerosity (upper and lower rows)	numerosity (upper and lower rows)
Irrelevant Dimension	digit identity	digit identity	digit identity	digit identity
Interference	5 → "three" (digit identity to numerosity)	2 → "three" (digit identity to numerosity)	5 → "four" (digit identity to numerosity)	2 → "four" (digit identity to numerosity)
Suppression of Irrelevant Information	5	2	5	2
Response Suppression	"five"	"two"	"match" key	"match" key
Correct Response Preparation	"three"	"three"	"mismatch" key	"mismatch" key
Response Competition	between "five" and "three"	between "three" and "two"	between "match" and "mismatch" keys	between "match" and "mismatch" keys

6.3 Numerosity Matching Stroop Paradigm

The current experiment employed a newly devised numerosity matching paradigm in which subjects had to judge whether two display sets had the same numerosity (number of items). Digits 1-5 and numerosities 1-5 were used, following Pavese and Umiltà's (1998) small-numerosity condition of their Stroop experiment.

In Pavese and Umiltà's (1998) enumeration Stroop experiment, subjects had to make a verbal response to the numerosity of the display set (e.g., to say "three" to a display such as "2 2 2"). The Stroop effect manifested as interference (significant increase in mean reaction time across neutral and incongruent close trials) and facilitation (significant increase in mean reaction time across congruent and neutral and trials) in the small-numerosity condition, where numerosity sets never exceeded 5 items. In this condition, a distance effect was also observed – mean reaction time for incongruent close trials (incongruent trials where the identity of the digits was greater or smaller than the numerosity of the display set by 1, e.g., "2 2 2" or "4 4 4") was significantly slower than that for incongruent far trials (incongruent trials where the identity of the digits was greater or smaller than the numerosity of the display set by 2, e.g., "1 1 1" or "5 5 5").

Furthermore, Pavese and Umiltà (1998), in the small-numerosity condition, found evidence for the two predictions based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970) which assumes that the distance between adjacent magnitude representations associated with numbers on the mental number line decrease as the number increases: 1) the effect of numerosity – "for a given arithmetic difference between digit identity and enumeration response, interference should increase with numerosity" (e.g., a stronger interference in the display 4 4 4 4 4 than in the display 4 4 4) and 2) the effect of digit identity (or numerical magnitude) – "interference should be greater when the digits to be counted are larger than the enumeration response than when they are smaller than the enumeration response" (e.g., a stronger interference in the display 5 5 5 5 than in the display 3 3 3 3).

Both dimensions – namely, numerical magnitude and numerosity – used in Experiments 6a and 6b require conceptual processing. According to Zorzi and Butterworth (1999), numbers evoke discrete magnitude representations. These representations would be expected to interact with the numerosity of the display set, which is also a discrete dimension. Experiments 6a and 6b aimed to replicate the Stroop effect (manifested as interference and facilitation). Following Pavese and Umiltà's (1998, 1999) experiments, an intra-stimulus distance effect was expected, such that more interference would be observed when the numerical magnitude and numerosity are close then when they are distant. Such an effect would indicate refined processing of numerical magnitude. Furthermore, an inter-stimulus distance effect was also predicted, i.e., a reduction in reaction time with increasing difference between the numerosity sets. The two predictions based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970) would also be examined. In addition, two programs – one written in Cogent and the other in JavaScript – were used, and their equivalence was tested.

6.4 *Experiment 6a: Numerosity Matching Stroop Task with Presentation Time of 100 ms*

6.4.1 Methods

6.4.1.1 Task

The task required the subjects to decide whether the numerosity in the first display of each trial matched the numerosity of the second display. The subjects had to make a key press in each trial and their reaction times and responses were recorded.

The task was computer-based. Two programs were used: one written in Cogent (running on a MATLAB Version 6.1 platform) and the other in JavaScript. These were included so that program equivalence could be examined.

6.4.1.2 Stimuli

In each trial, two displays containing numbers (Arabic digits), neutral stimuli ("X"), or hashes ("#") appeared in white on a black background sequentially on a 15" TFT screen. Each presentation lasted 100 ms, and there was no interval between the two displays in any given trial. The first display appeared on the top half of the screen and second display appeared at the bottom half of the screen. There were 10 possible locations in each display and they were equally spaced on an imaginary circle of diameter approximately 5 cm. After the two displays had been shown, there was an interval of 2000 ms before the subsequent trial.

There were two interference directions: forward and backward. In the forward interference condition, the first display always contained either numbers or neutral stimuli ("X"), and the second display always contained hashes ("#"); whereas in the backward interference condition, the first display always contained hashes ("#"), and the second display always contained numbers or neutral stimuli ("X"). Both numbers and numerosities ranged from 1 to 5. Thus, there were five levels of inter-stimulus distance (difference between the numerosities: 0, 1, 2, 3, and 4). However, inter-stimulus distance was completely confounded by whether the trial was "match" or "mismatch"; when distance was 0, trials were always "match", but when distance was 1 or greater, trials were always "mismatch".

There were 3 levels of congruity: congruent, neutral, and incongruent. Congruity was manipulated by varying the difference between the numerosity of the display and the stimuli contained in the display. In neutral trials, the stimuli always contained X's. In congruent trials, the identity of the numbers was the same as the numerosity of the display (e.g., 3 3 3). In incongruent trials, the identity of the numbers was different from the numerosity of the display (e.g., 5 5 5). The incongruent trials further divided into four levels of intra-stimulus distance (e.g., the display 4 4 4 4 4 had an intra-stimulus distance of 1, whereas the display 1 1 1 1 1 had an intra-stimulus distance of 4).

Each subject performed three blocks of trials. The order of these blocks was counterbalanced. Each block consisted of 120 trials. There were 40 congruent, 40 neutral, and 40 incongruent trials in each block. Half the trials were "match" (i.e., the two displays had the same numerosity), and the other rest of the trials were "mismatch" (i.e., the two displays had different numerosities).

The trials followed a pseudorandom order with the following constraints (in each block): 1) trials were not "match" (or "mismatch") for more than 3 consecutive trials, 2) the correct response did not occur on the same side for more than 3 consecutive trials, 3) trials were not of the same congruity level for more than 2 consecutive trials, 4) the same stimuli (numbers) did not appear in consecutive trials, except neutral ones, and 5) the same numerosity did not appear in consecutive trials.

6.4.1.3 Subjects

48 Japanese subjects (24 males and 24 females), age ranged 18 to 36 (mean = 23.9 years, standard deviation = 4.5 years) took part. All the subjects had studied English as their second language at high school (which started at the age of 13) for at least 6 years. Half of the subjects were tested with the Cogent program, and half with the JavaScript program. Half of each group performed the forward interference task, and the rest performed the backward interference task. Half of the subjects pressed the "F" key (with their left index finger on a qwerty keyboard) for "match" trials and the "J" key (with their right index finger on a qwerty keyboard) for "mismatch" trials, whereas the other half did the reverse.

6.4.2 Results

ANOVAs were used to analyse the mean percentage errors and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

6.4.2.1 Analyses on Mean Error Rates

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 2000 ms after the stimulus offset.

A 3 x 2 x 2 mixed-design ANOVA was conducted on mean error rates. The within-subjects factor was congruity (congruent, neutral, and incongruent), and the between-subjects factors were program (Cogent and JavaScript) and interference direction (forward interference and backward interference). The ANOVA revealed a significant main effect of congruity ($F_{(2,88)} = 13.21$, $p < 0.001$), but all other main effects and interactions were non-significant (all *n.s.*). Tests of within-subjects contrasts revealed significant differences in mean error rates between congruent and neutral trials ($F_{(1,44)} = 24.15$, $p < 0.001$), and between neutral and incongruent trials ($F_{(1,44)} = 13.42$, $p \leq 0.001$), but a non-significant difference between congruent and incongruent trials ($F_{1,44)} < 1$, *n.s.*). The mean error rates were 7.9%, 9.0%, and 8.1% for congruent, neutral, and incongruent trials respectively and the overall mean error rate was 8.3%.

6.4.2.2 Analyses on Mean Reaction Times

Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were removed prior to reaction time analyses.

A 5 x 3 x 2 x 2 mixed-design ANOVA was conducted on mean reaction times. The within-subjects factors were inter-stimulus distance (numerosity difference: 0, 1, 2, 3, and 4), and congruity (congruent, neutral, and incongruent), and the between factor were program (Cogent and JavaScript), and interference direction (forward interference and backward interference). The ANOVA revealed a significant main effect of inter-stimulus distance ($F_{(3,119)} = 48.84$, $p < 0.001$), a significant inter-stimulus distance x congruity interaction ($F_{(6,256)} = 5.64$, $p < 0.001$), and a significant main effect of program ($F_{(1,44)} = 9.24$, $p < 0.005$). The mean reaction time for Cogent program users was 736 ms (s.d. = 98 ms) and that for JavaScript program users was 642 ms (s.d. = 132 ms). All other main effects and interactions were non-significant (all *n.s.*). The mean reaction times for congruent, neutral, and incongruent trials were 689 ms, 689 ms, and 691 ms respectively. Within-subjects contrasts revealed both a significant linear trend ($F_{(1,44)} = 28.29$, $p < 0.001$) and a significant quadratic trend ($F_{(1,44)} = 97.76$, $p < 0.001$) for the factor inter-stimulus distance (see Figure 6.1).

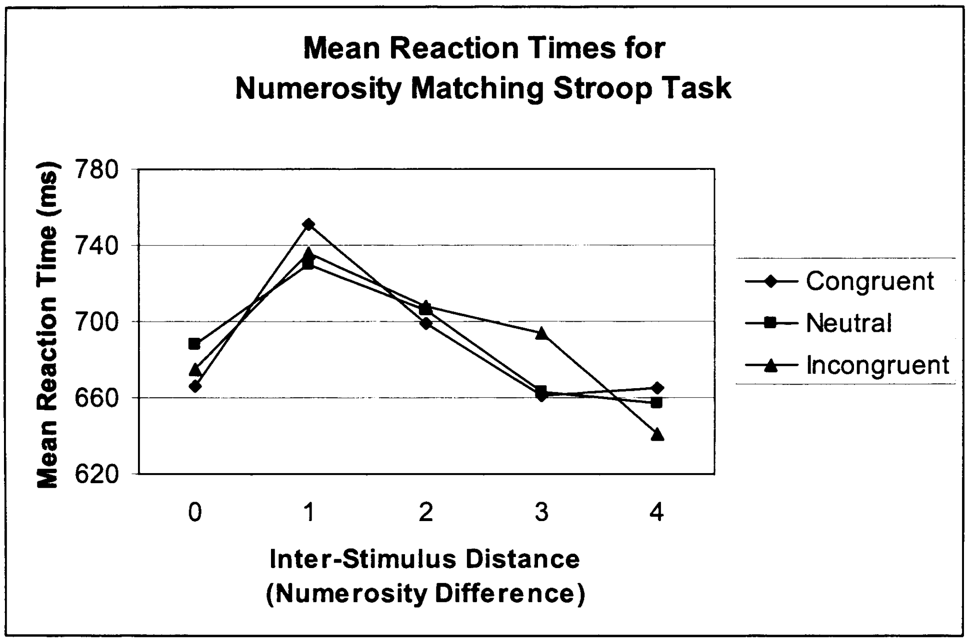


Figure 6.1 Mean reaction times (ms) for numerosity matching Stroop task at different levels of congruity across all subjects (Experiment 6a)

Figure 6.1 shows that mean reaction times peaked when inter-stimulus distance was 1 and there was a generally downward trend as distance increased from 1 to

4. Mean reaction times at distance 0 clearly did not follow the downward trend. Note that when the inter-stimulus distance was 0, the trials were always "match", whereas when the distance was not 0, the trials were "mismatch", and that there were more different "mismatch" trials than "match" trials (so even though there was an equal total number of "match" and "mismatch" trials, any given "match" trial was repeated more frequently than any given "mismatch" trial). The reaction time pattern therefore suggests that subjects might have employed different strategies for "match" and "mismatch" trials; given the more frequent repetition of "match" trials, subjects might have relied on a memory strategy for these trials (i.e., they might have responded upon memory of previous trials rather than upon comparison per se). Consequently, "match" trials were removed from further analyses.

Although the main effect of program was significant, this factor did not interact significantly with any other factor, so mean reaction times were collapsed over it. Similarly, the main effect of interference direction was non-significant and this factor did not interact with any other factor, so mean reaction times were collapsed over it.

A 4 x 3 repeated-design ANOVA was conducted on mean reaction times. The within-subjects factors were inter-stimulus distance (numerosity difference: 1, 2, 3, and 4) and congruity (congruent, neutral, and incongruent). It revealed a significant main effect of inter-stimulus distance ($F_{(2,87)} = 72.39$, $p < 0.001$), a non-significant main effect of congruity ($F_{(2,94)} < 1$, *n.s.*), and a significant inter-stimulus distance x congruity interaction ($F_{(5,227)} = 5.23$, $p < 0.001$). The mean reaction times for congruent, neutral, and incongruent trials were 694 ms, 689 ms, 695 ms respectively.

Within-subjects contrasts revealed both a significant linear trend ($F_{(1,47)} = 103.45$, $p < 0.001$) and a significant quadratic trend ($F_{(1,47)} = 7.04$, $p < 0.050$) for the factor inter-stimulus distance. At the linear level, the inter-stimulus distance x congruity interactions were non-significant (all *n.s.*). At the quadratic level, the inter-stimulus distance x congruity interactions were significant between congruent and neutral trials ($F_{(1,47)} = 5.67$, $p < 0.050$) and between neutral and incongruent trials ($F_{(1,47)} = 4.96$, $p < 0.050$).

Since the mean error rate was very high and the mean reaction times did not differentiate between the three levels of congruity, further analyses were not carried out. It was decided that a further experiment was to be conducted with modifications to the paradigm incorporated to improve performance accuracy (see Experiment 6b).

6.4.3 Discussion

The present study set out to investigate the autonomous processing of numerical magnitudes in a numerosity matching Stroop task, in which subjects had to judge whether two display sets had the same numerosity or not, while ignoring the identity of the stimuli. In addition, two programs – one written in Cogent and the other in JavaScript – were used and their equivalence was tested.

6.4.3.1 Program Equivalence

Although the mean reaction time of JavaScript program users was significantly faster than that of Cogent program users, the two groups did not differ significantly on mean error rates and the factor program did not interact significantly with any other factor in both error and reaction time analyses, suggesting that patterns of mean error rates and mean reaction times were similar in the two groups.

The significant difference in mean reaction times between Cogent and JavaScript users could be attributed to individual differences – subjects who were tested using the Cogent program had not only a longer mean reaction time of 736 ms but also a smaller standard deviation of 98 ms, compared to JavaScript program users' mean of 642 ms and standard deviation of 132 ms. There was no reason to suggest any inherent delay in time recording with Cogent, giving rise to a systematic error.

6.4.3.2 The Stroop Effect

The Stroop effect – both interference and facilitation – was not observed with regard to mean reaction times. More intriguingly, mean error rate was significantly higher in neutral trials than in incongruent trials. At first glance, an absence of the Stroop effect appears to suggest that the introduction of a matching process might have prevented the interference process. However, the high difficulty level, reflected by the higher mean error rate during neutral trials compared to incongruent trials, could also have contributed the absence of the Stroop effect. The present task, which required subjects to compare the numerosities of two displays, was more complicated than that used by Pavese and Umiltà (1998), which simply required subjects to enumerate the numerosity of a display in each trial.

In the current experiment, the high mean error rate for neutral trials may be explained by similarity of stimuli used in these trials – subjects saw neutral stimuli "X" and hashes "#" which are perceptually very similar, and this might have confused the subjects. Since, in each trial, the presentation time for the two displays was very brief (100 ms each) and there was no delay between them, having perceptually similar stimuli in the two displays might have made the task more difficult.

Since the mean error rate of the current experiment was high and significantly more errors were committed in neutral trials compared to congruent trials, the credibility of the current findings was questionable. Consequently a further experiment was carried out in attempt to validate the current findings (see Experiment 6b).

6.5 *Experiment 6b: Numerosity Matching Stroop Task with Presentation Time of 150 ms*

The current experiment employed the same experimental paradigm as in Experiment 6a. Modifications were made, primarily to improve performance accuracy. Additional analyses would be carried out examining (1) the processing of the task-irrelevant dimension – numerical magnitude, and (2) predictions based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970).

6.5.1 Method

6.5.1.1 Task

The task was exactly the same as that in Experiment 6a, where subjects had to decide whether the numerosity in the first display of each trial matched the numerosity of the second display. The subjects had to make a key press in each trial and their reaction times and responses were recorded. In the current experiment, there was only one interference direction, namely forward. The backward one was excluded since no difference was observed between the two in Experiment 6a.

A program written in Cogent (running on a MATLAB Version 6.1 platform) was used.

6.5.1.2 Stimuli

A few modifications were incorporated in the current experiment. Firstly, the direction of interference was restricted to forward only. Secondly, the symbol "@" instead of "X" was used as neutral stimuli in order to avoid any confusion which might have arisen in Experiment 6a between "X" and "#". Thirdly, the presentation time changed from 100 ms to 150 ms for both displays in an attempt to improve performance accuracy. All other procedures remained the same as Experiment 6a.

6.5.1.3 Subjects

28 subjects (11 males and 17 females), age ranged 17 to 36 (mean = 24.9 years, standard deviation = 5.1 years) took part. All the subjects were fluent in English. Half of the subjects pressed the "F" key (with their left index finger on a qwerty keyboard) for "match" trials and the "J" key (with their right index finger on a qwerty keyboard) for "mismatch" trials, whereas the other half did the reverse.

6.5.2 Results

ANOVAs were used to analyse the mean percentage errors and mean reaction times, and whenever Mauchly's test of sphericity assumption was violated, the Greenhouse-Geisser Epsilon was used to correct the degrees of freedom.

6.5.2.1 Analyses on Mean Error Rates

Errors included (1) incorrect responses made in the comparison tasks, i.e., subjects pressed the wrong key, and (2) trials where subjects failed to make a key press within the first 2000 ms after the stimulus offset.

A one-way ANOVA was conducted on mean error rates. The within-subjects factor was congruity (congruent, neutral, and incongruent). The ANOVA revealed a significant main effect of congruity ($F_{(2,54)} = 14.53, p < 0.001$). Tests of within-subjects contrasts revealed a significant difference in mean error rates between incongruent and neutral trials ($F_{(1,27)} = 12.60, p \leq 0.001$) and a non-significant difference between the latter and congruent trials ($F_{(1,27)} = 3.37$, *n.s.*). The mean error rates were 5.4%, 4.3%, and 3.9% respectively and the overall mean error rate was 4.5%. The overall mean error rate was significantly lower in the present experiment than in Experiment 6a ($t_{(74)} = 6.79, p < 0.005$).

6.5.2.2 Analyses on Mean Reaction Times

Reaction time outliers – values that were more than 1.5 x the interquartile range above the third quartile or 1.5 x the interquartile range below the first quartile – were removed prior to reaction time analyses.

A 5 x 3 repeated-measures ANOVA was conducted on mean reaction times. The within-subjects factors were inter-stimulus distance (numerosity difference: 0, 1, 2, 3, and 4) and congruity (congruent, neutral, and incongruent). The ANOVA revealed a significant main effect of inter-stimulus distance ($F_{(2,65)} = 31.53$, $p < 0.001$), a significant main effect of congruity ($F_{(2,54)} = 9.10$, $p < 0.001$), and a non-significant inter-stimulus distance x congruity interaction ($F_{(5,136)} = 1.21$, *n.s.*). Within-subjects contrasts revealed significant differences in mean reaction times between congruent and neutral trials ($F_{(1,27)} = 4.20$, $p \leq 0.050$), and between the latter and incongruent trials ($F_{(1,27)} = 4.96$, $p < 0.050$). The mean reaction times for congruent, neutral, and incongruent trials were 684 ms, 690 ms, and 703 ms respectively. Within-subjects contrasts also revealed a significant linear trend ($F_{(1,27)} = 33.17$, $p < 0.001$) and a significant quadratic trend ($F_{(1,27)} = 20.18$, $p < 0.001$) for the factor inter-stimulus distance (see Figure 6.2).

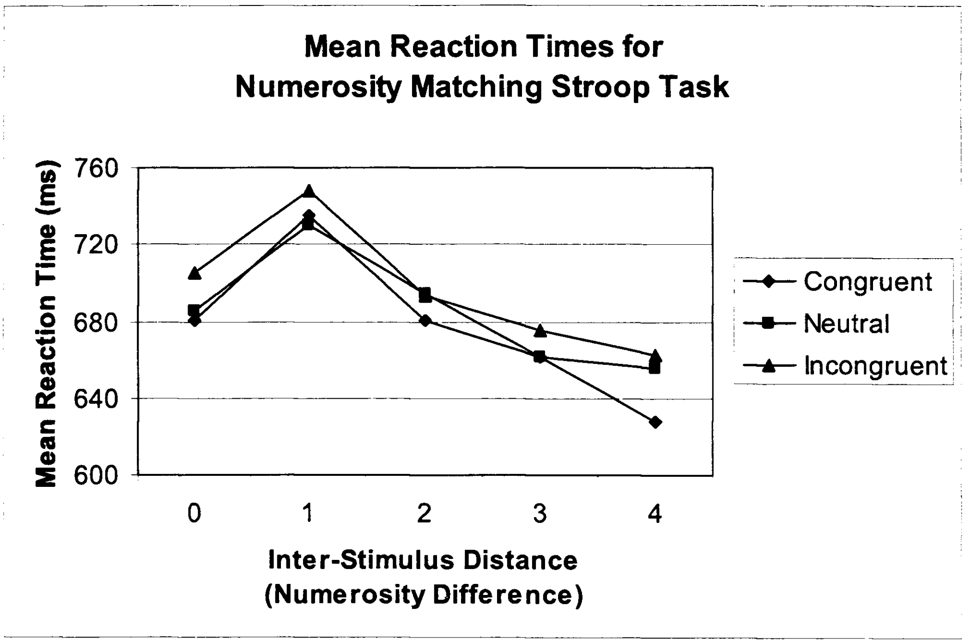


Figure 6.2 Mean reaction times (ms) for numerosity matching Stroop task at different levels of congruity across all subjects (Experiment 6b)

Figure 6.2, closely resembling to Figure 6.1, shows that mean reaction times peaked when the inter-stimulus distance was 1 and there was a generally downward trend as distance increased from 1 to 4. As in Figure 6.1, mean reaction times at distance 0 clearly did not follow the downward trend. When the inter-stimulus distance was 0, the trials were always "match", whereas when the distance was not 0, the trials were "mismatch", and the former were removed from further analyses (see Section 6.4.2.2 for details).

A 4 x 3 repeated-design ANOVA was conducted on mean reaction times. The within-subjects factors were inter-stimulus distance (numerosity difference: 1, 2, 3, and 4) and congruity (congruent, neutral, and incongruent). It revealed a significant main effect of congruity ($F_{(2,54)} = 7.60, p \leq 0.001.$). Tests of within-subjects contrasts revealed a significant difference in mean reaction times between congruent and neutral trials ($F_{(1,27)} = 5.20$, $p < 0.050$) and a non-significant difference between the latter and incongruent trials ($F_{(1,27)} = 3.22$, *n.s.*). The mean reaction times were 676 ms, 686 ms, and 695 ms respectively. The ANOVA also revealed a significant main effect of inter-stimulus distance ($F_{(2,57)} = 63.49$, $p < 0.001$) and a non-significant inter-stimulus distance x congruity interaction ($F_{(6,162)} = 1.19$, *n.s.*). Within-subjects contrasts revealed both a significant linear trend ($F_{(1,27)} = 98.50$, $p < 0.001$) and a significant quadratic trend ($F_{(1,27)} = 14.77$, $p \leq 0.001$) for the factor inter-stimulus distance.

Two additional analyses were carried out. These allowed predictions formulated by Pavese and Umiltà (1998) based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970) to be tested. Only "match" trials were selected to control for influence from the second display.

Firstly, the compressed number line hypothesis (Dehaene, 1992; Restle, 1970), predicting that, for a given intra-stimulus distance, interference should be greater with increasing numerosity, was tested. A 2 x 2 within-subjects ANOVA was used, where the factors were intra-stimulus distance (1 and 2) and numerosity (small and large, which referred to 1 or 2 unit(s) smaller and larger than the corresponding magnitude respectively). Table 6.6 shows the 2 x 2 design.

Table 6.6 The 2 x 2 design used to test for a numerosity effect in the numerosity matching Stroop task (Experiment 6b)

Intra-Stimulus Distance	Numerosity	
	Small	Large
1	2 3 3 4 4 4	2 2 2 3 3 3 3 4 4 4 4 4
2	3	3 3 3 3 3

The ANOVA revealed a significant main effect of numerosity ($F_{(1,25)} = 37.45$, $p < 0.001$), a non-significant main effect of intra-stimulus distance ($F_{(1,25)} < 1$, *n.s.*), and a non-significant numerosity x intra-stimulus distance interaction ($F_{(1,25)} < 1$, *n.s.*). The mean reaction times for small and large numerosities were 641 ms and 776 ms respectively.

Secondly, the compressed number line hypothesis (Dehaene, 1992; Restle, 1970) predicted that, for a given numerosity, interference should be greater when the numerical magnitude of the digits presented was larger than the numerosity of the display compared to a magnitude which was smaller than the numerosity. A 2 x 3 within-subjects ANOVA was used, where the factors were numerical magnitude (1 unit smaller and 1 unit larger than the corresponding numerosity) and numerosity (2, 3, and 4). The intra-stimulus distance was kept constant at 1. Table 6.7 shows the 2 x 3 design.

Table 6.7 The 2 x 3 design used to test for a numerical magnitude effect in the numerosity matching Stroop task (Experiment 6b)

Numerical Magnitude	Numerosity		
	2	3	4
Small	1 1	2 2 2	3 3 3 3
Large	3 3	4 4 4	5 5 5 5

The ANOVA revealed a significant main effect of numerosity ($F_{(2,35)} = 20.02$, $p < 0.001$), a non-significant main effect of magnitude ($F_{(1,22)} < 1$, *n.s.*), and a non-significant numerosity x numerical magnitude interaction ($F_{(2,44)} = 1.80$, *n.s.*). The mean reaction times for small and large numerical magnitudes were 720 ms and 715 ms respectively.

6.5.3 Discussion

The present study set out to investigate the autonomous processing of numerical magnitudes in a numerosity matching Stroop task, in which subjects had to judge whether two display sets had the same numerosity or not, while ignoring the identity of the stimuli. Predictions based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970) were also examined.

By modifying the paradigm used in Experiment 6a, an improvement in accuracy was observed in the current experiment (mean error rates were 8.3% and 4.5% respectively). In addition, the Stroop effect manifested as both interference and facilitation in the present experiment, whereas only the latter was observed in Experiment 6a.

6.5.3.1 The Stroop Effect, Distance Effects, and Autonomous Processing

Unlike Experiment 6a, evidence for the Stroop effect was observed with regard to both mean error rates and mean reaction times in the current experiment. Interference was reflected by the significantly higher mean error rate in incongruent trials compared to neutral trials, and facilitation by the significantly faster mean reaction time in neutral trials compared to congruent trials.

In the present experiment, an inter-stimulus distance effect was also observed as predicted, i.e., a reduction in mean reaction times as the inter-stimulus distance increased. For instance, it would take longer to process 5 items against 4 than to process 5 items against 1. Since numerosity was the task-relevant dimension, the inter-stimulus distance effect was in tune with the classic symbolic distance effect observed in previous research.

Two further analyses were carried out in order to test the predictions formulated by Pavese and Umiltà (1998) based on the compressed number line hypothesis (Dehaene, 1992; Restle, 1970). Before discussing the findings in relation to the predictions based on compressed number line, it is important to point out that the first of these analyses provided another opportunity to examine the autonomous processing of numerical magnitudes.

In the analysis examining the effect of numerosity predicted based on the compressed number line hypothesis, mean reaction times were computed over a restricted set of trials; the trials included were carefully selected considering the relationship between numerical magnitude and numerosity (see Table 6.6). (Note that it was impossible to include more levels of intra-stimulus distance when controlling for numerosity, since the number of possible different trials decreases as the intra-stimulus distance increases. For example, with an intra-stimulus distance of 3, the small numerosity condition would be 4 and the large numerosity condition would be 4 4 4 4 4 4 4; the latter exceeded the range of the numerosities used (1-5) in the current design.) The present analysis revealed a non-significant intra-stimulus distance effect, thus providing no evidence to support autonomous processing of numerical magnitudes during numerosity matching. This is inconsistent with the finding of Pavese and Umiltà (1998) who reported that more interference was generated by a display where the difference between the identity of the numbers and the numerosity was small (e.g., 4 4 4 4 4) than by one where the difference was large (e.g., 1 1 1 1 1).

The absence of an intra-stimulus distance effect might however be explained by the limited number of levels used. Since only two levels of intra-stimulus distance (1 and 2) were employed, polynomial contrasts could not be established; in other words, only a mere difference could be tested between close and distant conditions. Within the comparison Stroop paradigm, the use of limited distance levels (e.g., 2) often reduced the chance of the emergence of a distance effect, whilst the incorporation of more levels (e.g., in Experiments 2 and 3b of the present thesis) would allow potential distance effects to emerge. (See Section 3.2.4.1 for a discussion on the impact of the use of only two levels on the distance effect.)

In Pavese and Umiltà's (1998) experiment, the presentation time of the stimulus array was unlimited, i.e., until a response was made. Such a procedure might have allowed subjects to process the task-irrelevant numerical magnitude at a later stage, i.e., after the enumeration process has been completed. However, even with a fixed presentation time of 200 ms, Pavese and Umiltà (1999) managed to replicate the numerical distance effect. The absence of such an effect in the present experiment might be attributed to the high difficulty level; the current task, which required subjects to compare the numerosities of two displays, was more complicated than that used by Pavese and Umiltà (1998), which simply required subjects to enumerate the numerosity of a display in each trial. In addition, a shorter presentation time of 150 ms (for each of the two displays in each trial) was used, further contributing to difficulty level of the present task.

The Stroop effect (interference reflected by both reaction times and error rates) suggests that numerical magnitudes did influence (interfere) numerosity processing, but the absence of an intra-stimulus distance effect suggests that numerical magnitude processing was coarse. However, as discussed above, this absent effect might be attributed to the experimental design and the difficulty level. Although, currents findings are consistent with Tzelgov et al.'s (1992) suggestion that "only a crude, dichotomous representation of numerical size was encoded", the possibility that an intra-stimulus distance effect could emerge if a larger range of stimuli were employed (so that more levels of intra-stimulus distance could be used in the analysis) should not be excluded.

6.5.3.2 Effects of Numerosity and Numerical Magnitude

In the same analysis, however, a significant effect of numerosity was observed, replicating Pavese and Umiltà's (1998) finding. In other words, for a given intra-stimulus distance, interference was stronger for larger numerosities than for smaller ones (reflected by a significantly slower mean reaction time for the former, e.g., 4 4 4 4 4 vs. 4 4 4). This finding may be taken as support for the compressive number line hypothesis (Dehaene, 1992; Restle, 1970), according to which the mental distance between 5 and 4 would be smaller than that between 3 and 4.

However, the effect of numerosity may be explained in terms of the larger number of possible patterns that may be formed with larger numerosities (Peterson & Simon, 2000). Although the reaction times for numerosity enumeration within the subitizing range (6 or fewer according to Kaufman et al., 1949) increase in a less pronounced fashion with increasing numerosity than those beyond the range, Pavese and Umiltà (1998) reported a significant linear trend of increasing reaction time with increasing numerosity in the range 1 to 5 (with an average increment of 25 ms), where all adjacent contrasts were significant except between displays of 4 and 5 items. Thus, the current finding – the effect of numerosity – may be interpreted as a mere increase in difficulty in subitizing larger numerosities due to the larger number of possible patterns that may be formed with them.

In the final analysis, the effect of numerical magnitude was examined, but no evidence for such an effect was observed. In other words, for a given numerosity, mean reaction times did not differ significantly when the numerical magnitude of the digits presented was larger than the numerosity of the display compared to a numerical magnitude which was smaller than the numerosity (e.g., 5 5 5 5 vs. 3 3 3 3). The finding is therefore inconsistent with Pavese and Umiltà's (1998) prediction based on the compressive number line hypothesis. Instead, the finding can be taken as support for linear representations of numbers (e.g., Verguts et al., 2005; Zorzi & Butterworth, 1999). According to these authors, numerical representations are linear (i.e., non-compressive), and the distance between numerical representations does not decrease with increasing numerical magnitude, thus the absence of a numerical magnitude effect is consistent with this view.

6.5.3.3 Summary

The present experiment set out to investigate the processing of the task-irrelevant dimension – numerical magnitude. Although the Stroop effect suggests that numerical magnitudes influenced the processing of numerosity information, the absence of an intra-stimulus distance effect suggests that numerical magnitudes were not processed in a refined fashion. This appears to be consistent with Tzelgov et al.'s (1992) suggestion that "only a crude, dichotomous representation of numerical size was encoded." However, the absence of an intra-stimulus distance might have arisen due to the limitations of the present experimental design – a restricted number of distances used.

Nevertheless, the current findings again highlight the limitations in using the Stroop effect to make inferences about autonomous information processing; the effect merely implies an influence from the task-irrelevant dimension, but fails to give an accurate indication as to the degree of information processing (i.e., whether information is processed in a refined fashion or not). As discussed in previous experiments of the present thesis (Experiments 2 and 3b), distance effects can be used reliably to indicate refined information processing. It is important to stress that three or more distances should be employed, so that polynomial contrasts can be specified in the analysis to test for linear (and any higher order) trends; the mere difference between distant and close pairs reflects some degree of processing, but cannot be reliably used to reflect refined information processing. Furthermore, a difference was more likely to emerge if the two distances were very different than if they were similar (see Section 3.2.4.1 for details). In addition, findings of the present experiment provide sound evidence for linear (non-compressive) numerical representations.

7 General Discussion

The series of experiments presented in this thesis aimed to investigate the effects of task relevance on numerical magnitude processing. Existing experimental paradigms have been modified and new paradigms devised to manipulate numerical magnitude as both the task-relevant and -irrelevant dimensions. Experimental designs were primarily Stroop variants or Stroop-like (Stroop, 1935). In all the experiments, stimuli were multidimensional and filtering of task-irrelevant information was involved. Tasks could be different in nature, e.g., numerical magnitude comparison task (which requires conceptual processing) and physical size comparison task (which can be solved perceptually), or similar in nature, e.g., numerical magnitude and parity (both are semantic aspects of numbers and require conceptual processing), and numerical magnitude and numerosity (both require conceptual processing; the latter can be viewed as the non-symbolic representation of the former). Other factors which could potentially influence conceptual magnitude processing were also examined, and they include writing system and familiarity.

The present thesis argues that numerical magnitude is processed autonomously. In other words, even when numerical magnitude is the task-irrelevant, one cannot avoid but process the information. Furthermore, this autonomous processing is refined in most cases. Refined numerical magnitude processing in number comparison and parity tasks refers to precise judgements on a graded scale, as opposed to a coarse large-small dichotomous classification of numbers. Numerical magnitudes, when task-irrelevant, did not simply influence the task-relevant comparisons (Experiments 2 and 3b) and parity judgements (Experiment 5a), but were processed with precision resulting in reversed numerical distance effects.

More importantly, the experimental designs and findings highlight the methodological flaws and possible misinterpretations of results in previous research. The current chapter addresses the following questions: how to measure semantic (in the present context, numerical magnitude) processing and how to systematically examine the degree of information processing in both task-relevant and -irrelevant dimensions.

7.1 Measuring Numerical Magnitude Processing

The comparison Stroop paradigm has traditionally been used to examine numerical magnitude processing. A typical experiment consists of two tasks, namely numerical magnitude comparison and physical size comparison tasks. A pair of numbers (differing in terms of numerical magnitude and/ or physical size) is presented to the subject, who is asked to judge which of the pair is larger (or smaller) either numerically or physically depending on the task requirement. The Stroop effect – manifested as both interference (an increase in reaction time and/ or error rate in incongruent trials relative to neutral trials) and facilitation (a decrease in reaction time and/ or error rate in congruent trials relative to neutral trials) – is interpreted to reflect autonomous processing of the task-irrelevant dimension. It has been reported that the Stroop effect is more prominent during numerical comparisons than during physical comparisons (Henik & Tzelgov, 1982). In other words, physical size, as a task-irrelevant dimension, exerts more influence on the task-relevant dimension, numerical magnitude, but the degree of such influence was never systematically measured. Henik and Tzelgov's (1982) pointed out this precise problem, "It is impossible, at this point, to claim that the numerical information is fully processed."

Attempts have been made to examine the degree of numerical magnitude information processing during physical comparisons by varying the numerical distance (or difference) between the two numbers presented and a reversed numerical distance effect – an increase in reaction time as the numerical distance between the two numbers increases – has been reported during physical comparisons (Girelli et al., 2000; Henik & Tzelgov, 1982). This contrasts with the classic numerical distance effect – a decrease in reaction time as the numerical distance between the two numbers increases – observed during numerical comparisons. The reversed numerical distance effect had received very little attention from researchers, possibly due to difficulties in replicating (see Rubinsten et al., 2002 who failed to replicate the effect). However, the inconsistency may be explained in terms of poor experimental design. No explicit attempt had ever been made to match the two dimensions (namely numerical magnitude and physical size), and this could impact directly on the pattern of reaction time in comparison performance.

Matching the two dimensions, physical size and numerical magnitude, is not straightforward. Firstly, these dimensions are of very different natures – physical size comparisons can be solved perceptually, whereas numerical magnitude comparisons require conceptual processing. Such a difference may actually explain the faster mean reaction time for physical size comparisons relative to numerical magnitude comparisons (Henik & Tzelgov, 1982; Tzelgov et al., 1992; see also Experiment 2 of the present thesis). It is practically impossible to match these dimensions in terms of reaction time, so attempts were made to match them in terms of error rate (see Experiments 3a and 3b). Furthermore, previous experiments (except Rubinsten et al., 2002) typically employed factorial designs, where nine numbers have been put into conflict with, at most, three sizes – usually a large size and a small size to construct congruent and incongruent trials, and a medium size for neutral trials, hence often only one level of physical distance was used. Thus, limited inferences could be made regarding the amount of information available in the task-relevant and -irrelevant dimensions. Thus, physical comparisons appeared to involve a mere large-small dichotomous classification rather than judgements on a graded scale. In the present series of experiments, attempts were made to develop a fully parametric design, where the two dimensions have the same number of levels (e.g., 6 numbers and 6 physical sizes in Experiment 2; 9 numbers and 9 physical sizes in Experiments 3a and 3b), and an equal number of distances was constructed in the two dimensions (e.g., 3 levels of numerical distance and 3 levels of physical distance in Experiments 2; 4 levels of numerical distance and 4 levels of physical distance in Experiments 3a and 3b).

When the two dimensions were properly matched, it was possible to examine the degree of information processing in both task-relevant and -irrelevant dimensions. This was done by using distance effects as indexes for refined information processing. Here, refined information processing refers to a precise comparison process between two numbers on a given dimension (numerical or physical), resulting in a (numerical or physical) difference between the numbers. This contrasts with the large-small dichotomous classification reflected by the Stroop effect. The following section directly compares inferences made based on these effects.

7.1.1 Stroop and Distance Effects when Comparing Arabic Digits

The manifestations of the Stroop and distance effects when comparing Arabic digits are summarised in Table 7.1. In Experiment 1 where a factorial design was employed, the Stroop effect was observed as interference and facilitation in the numerical task, but only as interference in the physical task. The latter, a perceptual-based task, was easier (significantly faster) to perform, and hence a lack of room for facilitation to be observed. This is consistent with the general finding in the Stroop literature that facilitation is virtually always substantially smaller than interference (see review, MacLeod, 1991).

Due to the limits of the design in Experiment 1, distance effects could only be tested for in the numerical dimension. The classic numerical distance effect was observed when numerical magnitude was the task-relevant dimension (i.e., during numerical comparisons). On the other hand, when numerical magnitude was the task-irrelevant dimension (i.e., during physical comparisons), numerically distant and close pairs did not differ significantly in terms of mean reaction time, but it is important to note that the mean reaction times were in the predicted direction, i.e., in the direction of a reversed numerical distance effect.

Experiment 2 was a first attempt to employ a parametric design – 3 levels of distance in each of two dimensions were used. The Stroop effect – manifested as interference and facilitation – was only observed in the numerical task, but not the physical task. The task-relevant dimension showed a classic distance effect in each of the two tasks, i.e., a numerical distance effect in the numerical task and a physical distance effect in the physical task. Furthermore, the task-irrelevant dimension showed a reversed distance effect in each of the two tasks, i.e., a reversed physical distance effect in the numerical task and a reversed numerical distance effect in the physical task. Such findings are strong evidence for refined autonomous processing of numerical magnitude and physical size information. In addition, these findings have highlighted the lack of sensitivity of the Stroop effect as a measure of refined information processing; for example, to conclude that numerical information did not influence physical size processing when the former was task-irrelevant based on an absence of the Stroop effect would be utterly inaccurate when there was a significant reversed numerical distance effect providing strong evidence for refined autonomous numerical magnitude processing. Thus, distance effects should always be used in the current comparison type of the Stroop paradigm as an indicator of refined information processing.

In Experiment 2, only 6 numbers (1, 3, 4, 6, 7, and 8) and 6 physical sizes were used to create 3 levels of distance in each of two dimensions. This was done to exclude inconsistencies in the Kana conditions (for details, see Section 2.4.1.2). On the other hand, in Experiments 3a and 3b, only Arabic digits were used as stimuli. Thus, a full range of 9 numbers (1-9) were used as well as 9 physical sizes to create 4 levels of distance in each of two dimensions. In Experiment 3a, the Stroop effect – manifested as interference and facilitation – was observed in both tasks. The task-relevant dimension showed a classic distance effect in each of the two tasks, i.e., a numerical distance effect in the numerical task and a physical distance effect in the physical task. However, the task-irrelevant dimension showed a reversed distance effect only in the numerical task, i.e., a reversed physical distance effect in the numerical task. The design of Experiment 3b was identical to that of Experiment 3a, except with a slight modification of increasing the absolute sizes of stimuli. Here, the Stroop effect – manifested as interference and facilitation – was observed in both tasks. The task-relevant dimension showed a classic distance effect in each of the two tasks, i.e., a numerical distance effect in the numerical task and a physical distance effect in the physical task. Furthermore, the task-irrelevant dimension showed a reversed distance effect in each of the two tasks, i.e., a reversed physical distance effect in the numerical task and a reversed numerical distance effect in the physical task. The distance effects replicated those in Experiment 2.

A fully parametric design of the comparison type Stroop paradigm, with appropriately matched dimensions, was finally developed in Experiment 3b. The Stroop effect, the classic distance effects in the task-relevant dimensions, and the reversed distance effects in the task-irrelevant dimensions, were demonstrated in both comparison tasks. By incorporating the full range of single digits (1-9) and 9 physical sizes, 4 levels of distance (instead of only 3 in Experiment 2) were constructed and used in Experiment 3b, and the distance effects observed (see Figure 3.5 and Figure 3.6) were more prominent than those in Experiment 2 (see Figure 2.17 and Figure 2.18).

The fact that the reversed numerical distance effect was not always observed or even examined in previous experiments could be attributed to poor experimental design. Two levels of numerical distance were often employed, giving rise to numerically distant and close trials, but if these conditions were similar (i.e., the two numerical distances were not highly discriminable), it would be difficult for a distance effect to emerge, especially under task-irrelevant condition. To illustrate this point further, one could examine Figure 3.6; if numerical distances 3 and 4 were chosen, a reversed numerical distance effect would be unlikely to emerge). Moreover, the use of two levels of distance meant that only a mere difference could be established between the distant and the close conditions, but not a linear trend per se. In contrast, Experiments 2 and 3b employed more levels of distance, allowing polynomial contrasts to be specified in the analyses, thus linear (and any other higher order) trends could be established.

The above series of experiments has, not only demonstrated the lack of sensitivity of the Stroop effect as a measure of refined information processing, but also highlighted the design problem – unbalanced dimensions – used in earlier experiments. The largely neglected reversed numerical distance effect (first observed by Henik & Tzelgov, 1982; replicated by Girelli et al., 2000) has proven to be an important index for refined autonomous processing of numerical magnitude. The parametric design has also allowed the equivalent effect in the physical dimension to manifest. Both behavioural and neuroimaging have provided unambiguous evidence for autonomous processing of numerical magnitude.

Table 7.1 Summary of the manifestations of the Stroop and distance effects reflected by mean reaction times

Experiment	Numerical Task				Physical Task			
	Facilitation	Interference	Distance		Facilitation	Interference	Distance	
			Numerical (Classic)	Physical (Reversed)			Physical (Classic)	Numerical (Reversed)
1	✓	✓	✓	N/A	x	✓	N/A	x
2	✓	✓	✓	✓	x	x	✓	✓
3a	✓	✓	✓	✓	✓	✓	✓	x
3b	✓	✓	✓	✓	✓	✓	✓	✓

7.1.1.1 Parietal Activities and Autonomous Numerical Magnitude Processing

Using the functional Magnetic Resonance Imaging (fMRI) technique, Experiment 3b examined the involvement of the parietal lobes in the processing of numerical information. Contrary to previous research findings (Pinel et al., 2001, 2004), no evidence was found to suggest that parietal activation was parametrically modulated by either numerical or physical distance. However, a closer look at the previous research findings suggests that the numerical distance modulated parietal activation observed by Pinel and colleagues (Pinel et al., 2001, 2004) might not be as robust as the authors claimed.

In both studies (Pinel et al., 2001, 2004), the reported distance related parietal activation was merely a main effect across two or three (respectively) levels of numerical distance as opposed to a significant linear decrease in activation with increasing numerical distance. Furthermore, a very small number of subjects was used in Pinel et al.'s (2001) analysis. Consequently, both reports by Pinel and colleagues (Pinel et al., 2001, 2004) on the distance modulated parietal activation were questionable (see Section 3.2.4.2 for a detailed discussion).

In Experiment 3b, enhanced activation was observed in several parietal regions (in right inferior parietal lobule, right precuneus and left superior parietal lobule) when processing numerical distance relative to physical distance. Furthermore, the parietal regions, which showed enhanced activation processing numerical relative to physical distance, were not affected by task requirements during conflict situations. In other words, these parietal regions were equally active whether or not required by the task to process numerical magnitudes. The lack of difference in parietal activation level across numerical and physical comparison tasks during conflict situations provides a strong piece of evidence for autonomous processing of numerical magnitude.

7.1.1.2 Effects of Writing System

The effects of writing system on the comparison type Stroop task performance were investigated by testing bilingual subjects – a group of Chinese-English bilinguals (Experiment 1) and a group of Japanese subjects who had studied English at high school (Experiment 2). It is the latter experiment where the first parametric design for the comparison Stroop paradigm was introduced and this is the one that the current discussion will be focused on.

In Experiment 1, a group of Chinese-English bilinguals performed the two comparison tasks with three types of stimulus, namely Arabic digits, Chinese and English written verbal numerals. Evidence for the Stroop effect was observed in both comparison tasks with Chinese written verbal numerals. Like Arabic digits, the Stroop effect observed with Chinese written verbal numerals during physical comparisons reflects autonomous processing of numerical magnitude. The same also applies to English written verbal numerals, but only with the control subjects. The reversed numerical distance effect was not observed throughout, thus providing no evidence for refined processing of numerical magnitude when it was task-irrelevant. This finding is not surprising given that an unbalanced factorial design was employed.

In Experiment 2, two Japanese writing scripts, Kanji (ideographic) and Kana (syllabic), were investigated. Japanese verbal numerals are usually written in Kanji and rarely in Kana. In this experiment, subjects performed the two comparison tasks with four types of stimulus, namely, Arabic digits, Kanji written verbal numerals and their unfamiliar Kana equivalents, and English written verbal numerals. Some (although limited) evidence for the Stroop effect was observed in both comparison tasks with Kanji numerals and their Kana equivalents. This is consistent with Foltz et al. (1984) who observed the Stroop effect not only with Arabic digits, but also with written verbal numerals. However, in Experiment 2, the Stroop effect was absent with English written verbal numerals. Such an absence may be explained in terms of subjects' lack of fluency in this language. The use of a fixed-pairs design (Foltz et al., 1984; see also Section 1.5.1) could have diluted any potential effect, and in a dysfluent language, the dilution of effect might have been more prominent.

Although evidence for the Stroop effect was observed with Kanji and Kana stimuli, closer inspections of the results revealed an absence of reversed distance effect during physical comparisons (i.e., when numerical magnitudes were task-irrelevant). This indicates that task-irrelevant information processing of these stimuli was coarse, unlike Arabic digits, the processing of which was refined regardless of task relevance.

In summary, numerical magnitudes are autonomously activated across writing systems (indicated by the Stroop effects). The difference rests on the degree of processing, i.e., whether the information is processed in a refined or coarse fashion various across writing systems. Distance effects provide an adequate indicator for refined information processing. These effects should always be reported alongside the Stroop effects as the latter have been proven insensitive as a measure for refined information processing. Factors such as fluency in the language of the stimuli and the strength of pathway to semantics (see Section 2.4.3.3) play important roles in determining the degree of information processing of numerical magnitudes.

7.1.1.3 Effects of Familiarity

Experiment 4 employed a variant of the comparison type Stroop paradigm to investigate the effects of familiarity when processing the value and the size of coins, which subjects would have daily encounters with. Monochromatic images of British coins were presented in their usual physical size ratios and in novel ratios. Subjects had to judge the larger of a pair either conceptually (coin value comparison) or physically (coin size comparison). The Stroop effect was observed in both tasks.

Interestingly, familiarity of the stimulus ratios had differential effects on conflict (incongruent) and non-conflict (congruent) trials during conceptual comparisons. When the physical size was congruent to the coin value, familiar pairs had a significant advantage over unfamiliar pairs during conceptual comparisons. Such a difference was, however, not observed with incongruent pairs, indicating that familiar conflicts were not overcome by subjects' extensive experience with the British coins. One possible explanation for the differential familiarity effects is that there are fewer "naturally-occurring" (familiar) conflict pairs than non-conflict pairs of coins, thus experience with the conflicting situations was relatively less (see Section 4.2.3 for a detailed discussion). This might have contributed to the lack of benefit produced by the familiar conflict trials.

The experiment has provided a new way to investigate numerical information processing with stimuli which had high ecological validity. It has also shed new light onto the complicated effects of familiarity (or practice) on the Stroop phenomenon.

7.1.2 Numerical Magnitude Processing during Parity Judgements

The comparison type Stroop paradigm employs two dimensions (namely, numerical magnitude and physical size) which are different in nature. Consequentially, the two comparison tasks might occur on very different levels – physical size comparisons can be solved perceptually whereas numerical magnitude comparisons require conceptual processing. Hence, other experiments have been devised with careful selection of dimensions. Firstly, Experiments 5a and 5b employed a parity judgement paradigm where both of the dimensions – parity and numerical magnitude – are semantic properties of numbers and require conceptual processing. Experiments 6a and 6b, which are further discussed in Section 7.1.3, employed a numerosity matching paradigm where both dimensions – numerosity and numerical magnitude – require conceptual processing; the former may be viewed as the non-symbolic representation of magnitude information.

In Experiment 5a, subjects had to decide in each trial whether two numbers (Arabic digits) had the same parity or not. The key finding is that the task-irrelevant dimension – numerical magnitude – showed a reversed numerical distance effect (which actually manifested as two effects or two positive linear trends: one for "same" parity trials and another for "different" parity trials), providing support for autonomous processing of numerical magnitude.

It may be argued that numerical magnitude processing is an obligatory part of the parity status decision since there is evidence that numerical magnitude information is more available than parity information (Sudevan & Taylor, 1987; see also Boles, 1986) and that the concept of parity is built upon the understanding of numerical magnitude (Miller & Gelman, 1983). Moreover, a neuropsychological dissociation has been reported in the direction of preserved numerical magnitudes and impaired parity information (Dehaene & Cohen, 1991) but a dissociation in the reversed direction has never been reported. Autonomous processing of numerical magnitude is therefore not surprising based on the above evidence. Limited evidence was also observed in Experiment 5b to support autonomous processing of numerical magnitude during parity judgements with English written verbal numerals.

On the other hand, the findings of Experiment 5a and 5b have also provided evidence that memorial strategies might have been in operation during parity judgements. It was observed that trials containing two even numbers were significantly faster than other trial types, supporting Dehaene et al.'s (1993) idea that the even series is better learnt (see also Dehaene & Cohen, 1991).

In summary, findings of Experiment 5a highlight the complexity of the parity judgement task and suggest that multiple strategies might be in place during complex parity judgements.

In addition, the stimulus-side effect – a prediction based on the SNARC effect that parity judgements would be faster with trials containing a small number on the left than on the right – was tested, but no significant difference in mean reaction times between these types of trials was observed in Experiments 5a and 5b. This finding does not support the idea of a left-to-right mental organisation of numbers.

7.1.3 Numerical Magnitude Processing during Numerosity Matching

In Experiment 6a and 6b, subjects had to decide whether two consecutive displays contained the same number of items or not, while ignoring the identities of the stimuli (e.g., 5 5 5 and # # # had the same numerosity). Some trials (congruent and incongruent) contained Arabic digits (e.g., 3 3 3 and 5 5 5 respectively), whilst other (neutral) contained non-numerical stimuli (e.g., @ @ @). Both dimensions – numerical magnitude and numerosity – required conceptual processing. The mean error rate in Experiment 6a was very high, thus modifications were made and Experiment 6b was conducted. This section will discuss the findings of the latter only.

In Experiment 6b, the Stroop effect (manifested as both interference and facilitation) was observed. This effect implies some level of autonomous numerical magnitude processing during numerosity matching. However, no evidence was found to support the idea that this autonomous processing was performed in a refined fashion. The experimental design and the high difficulty level of the present task might have contributed to the absence of refined numerical magnitude processing (see Section 6.5.3.1).

Despite failing to demonstrate refined autonomous processing of numerical magnitude during numerosity matching, findings of Experiment 6b have provided evidence for linear (non-compressive) numerical representations. Issues relating to the representations of numbers are discussed in Section 7.2.2.

Unlike the parity judgement paradigm used in Experiments 5a and 5b, numerical magnitude processing was undoubtedly task-irrelevant in the numerosity matching Stroop paradigm used in Experiments 6a and 6b. The uncertainty as to whether numerical magnitude processing plays an obligatory part in the decision of parity status has rendered the former paradigm less reliable than the latter as a task used for testing autonomous numerical magnitude processing.

7.1.4 Summary

The experiments presented in the current thesis converge to a consensus that the Stroop effect on its own is not a sensitive enough measure for refined autonomous processing of task-irrelevant information. In other words, the effect merely reflects some degree of autonomous processing, but provides no indication as to whether the information is processed in a refined fashion or not. Where available, distance effects should be reported alongside; the reversed distance effect, when observed in task-irrelevant conditions, indicates refined autonomous information processing.

The present findings highlight the importance of appropriate experimental designs. Whenever multidimensional stimuli are used, caution must be taken to select and match the dimensions. To investigate distance effects, parametric modulations should be introduced to each of the dimensions. In order to examine distance effects, at least three (ideally four) levels of distance should be employed. Appropriate analyses should also be adopted when examining distance effects; polynomial contrasts should be specified.

In addition to the Stroop and distance effects, the stimulus-side effect – a prediction based on the SNARC effect – was also examined, but such an effect was not observed. This finding does not support the idea of a left-to-right mental organisation of numbers.

7.2 *Towards a Comprehensive Theoretical Account of Numerical Processing and Representation*

7.2.1 Numerical Comparison

Experiments 1 to 4 employed the comparison type Stroop paradigm and from these experiments have come a consensus that numerical magnitudes are autonomously accessed regardless of task-relevance. The autonomous aspect is implicated by the Stroop effect. However, this effect does not provide any indicator as to whether or not information is processed in a refined fashion. Instead, the current thesis argues that distance effects provide an adequate measure for refined information processing.

Refined information processing was observed in both the task-relevant and -irrelevant dimensions during both numerical and physical comparisons (Experiments 2 and 3b). In the task-relevant dimension, refined information processing was marked by a linear decrease in reaction time in the task-relevant dimension, i.e., a classic numerical distance during numerical comparisons and a classic physical distance during physical comparisons, whereas in the task-irrelevant dimension, refined information processing was marked by a linear increase in reaction time, i.e., a reversed numerical distance during physical comparisons and a reversed physical distance during numerical comparisons. The latter effects indicate that the more discriminable the task-irrelevant stimuli the more influence they exerted on the comparison process of the task-relevant dimension.

When considering the Stroop effect, Foltz et al.'s (1984) relative speed of processing account provides a satisfactory explanation for the general finding that the Stroop effect was more commonly observed during the physical comparison task compared to the conceptual (numerical or coin value) comparison task. According to Foltz et al.'s (1984) account, the faster process (physical comparison) would exert more influence on the slower (conceptual comparison).

However, the reversed numerical distance effect observed during physical comparisons in Experiments 2 and 3b (see also Girelli et al., 2000; Henik & Tzelgov, 1982) challenges Foltz et al.'s (1984) account which predicted that physical size as a task-irrelevant dimension would exert a stronger influence on performance than numerical magnitude would. The reversed numerical distance effect indicates that the task-irrelevant numerical magnitudes were processed in a refined fashion. Foltz et al.'s (1984) account would be hard-pressed to explain such a strong influence exerted by the relatively slow numerical processing on the physical comparison task.

Tzelgov et al.'s (1992) account based on Logan's (1988) two-process theory of skilled performance (Section 1.5.2) provides an explanation for the classic numerical distance effect observed in the numerical task, but fails to explain the reversed numerical distance effect observed in the physical task.

Tzelgov et al.'s (1992) account explains comparison Stroop performance in terms of a race between memory-based and algorithm-based processes. The authors attributed the classic numerical distance effect in the numerical task to the algorithm-based process which allowed refined numerical magnitude processing. In three experiments, Tzelgov et al. (1992) failed to observe any numerical distance effect when numerical magnitude was the task-irrelevant dimension (Tzelgov et al., 1992, Experiments 1-3), and they attributed this absent effect to the dominance of the memory-based process which was used to perform a crude, dichotomous classification of numbers into large or small. However, the distinction between memory-based and algorithm-based processes failed to account for the refined information processing observed in both the numerical and the physical dimensions under both task-relevant and -irrelevant conditions in Experiments 2 and 3b (see also Girelli et al., 2000; Henik & Tzelgov, 1982).

The interaction between the two dimensions can be explained in terms of stimulus discriminability. Girelli (1998) pointed out that when the comparison was more demanding, less interference would be produced. For example, during a physical comparison, a numerically close pair would produce little interference, since it would be more effortful to perform the numerical comparison. This relates to Melara and Mounts' (1993) notion of discriminability. Pansky and Algom (1999) found that when the task-irrelevant dimension was more discriminable, a sizeable Stroop effect affected performance on the task-relevant dimension, but when it was less discriminable, the Stroop effect was considerably weaker. These findings echo the reversed distance effects observed in Experiments 2 and 3b (see also Girelli et al., 2000; Henik & Tzelgov, 1982), that distant pairs are more discriminable than close pairs, thus exerting a stronger influence on the task-relevant dimension.

The asymmetry that the physical dimension influences more strongly conceptual comparisons than conceptual dimension on physical comparisons can also be explained in terms of discriminability difference between the two dimensions. Physical size comparisons can be solved perceptually, whereas numerical magnitude and coin value comparisons require conceptual processing; thus it is reasonable to assume that the physical sizes are more easily discriminated than numerical magnitudes or coin values. Consequently, the former would have more influence on the latter.

An alternative account, which is not incompatible with the discriminability explanation, proposed by Cohen et al. (1990) explains, not only the Stroop effect (that it is more commonly observed in numerical tasks than in physical tasks), but also the differences in performance across writing scripts and effects of familiarity/ practice.

Cohen et al.'s (1990) parallel distributed processing model places an emphasis on the relative strengths (rather than speeds) of pathways in the processing system, assuming that the degree of automaticity is a function of the strength of each pathway; the relative speed of processing can be viewed as an indicator for the strength of pathway. Thus, the model explains the asymmetry that the Stroop effect is more commonly observed in numerical tasks than in physical tasks in terms of a stronger processing pathway in the physical dimension (the task-irrelevant dimension in the numerical task) compared to the numerical dimension (the task-irrelevant dimension in the physical task); thus the physical size as a task-irrelevant dimension would have more influence on the relevant task than numerical size as a task-irrelevant dimension. The difference in pathway strength is reflected by the faster processing time in physical tasks compared to numerical tasks. Moreover, Cohen et al.'s (1990) model also accounts for the asymmetry that interference is virtually always substantially smaller than facilitation by incorporating non-linear processing units in the model (see review, MacLeod, 1991).

Cohen et al.'s (1990) model provides an explanation for the differences in comparison Stroop performance across different writing scripts in bilingual subjects. In Experiments 1 and 2, two groups of bilingual subjects whose second language was English were tested. No Stroop effect was observed with these subjects in the physical task when the stimuli were English verbal numerals. However, both groups exhibited the Stroop effect in the physical task when the stimuli were written numerals in their corresponding first language. Within Cohen et al.'s (1990) framework, these differences would be explained in terms of relative strength of processing pathway. Numerals written in subjects' first language would undoubtedly have received more practice, hence the processing pathway would be stronger, and as a result, these stimuli were more likely to produce a Stroop effect than numerals written in the second language which would have received relatively less practice, hence a weaker pathway to semantics.

When discriminability is incorporated, as a factor influencing comparison Stroop task performance, into Cohen et al.'s (1990) model, the differential familiarity effects observed in Experiment 4 can be adequately explained. In Experiment 4, the Stroop effect was observed in both conceptual and physical comparison tasks with British coin images. During physical comparisons, familiarity had no effect on the Stroop effect. On the other hand, during conceptual comparisons, familiarity aided only the non-conflict (congruent) trials, but not conflict (incongruent) trials. Since there are more familiar congruent than incongruent coin pairs, the differential familiarity effects during conceptual tasks can be explained in terms of differential strengths of processing resulting from differential experience (stronger pathway from more frequent encounters with congruent than incongruent coin pairs), hence familiarity only aided congruent trials. The absence of a familiarity effect during physical comparisons may be explained in terms of an interaction between familiarity and stimulus discriminability; familiarity had no effect on the Stroop phenomenon when the task-irrelevant dimension – conceptual value – was relatively low in discriminability. It must be noted that the above explanations are speculative and further research is necessary to examine these suggestions.

Cohen et al.'s (1990) model has provided a satisfactory account of the findings reported in the present thesis. In addition, the model's assumption that automaticity is a function of the strength of each pathway is very important in understanding the developmental trend of the Stroop effect. Girelli et al. (2000) reported that the Stroop effect in the physical task did not emerge until the third grade and became highly significant in fifth grade. This finding has led to the implication that the autonomy of numerical magnitude processing relies on a firm understanding of number semantics and that practice gives rise to the strength of processing pathway.

In summary, Cohen et al.'s (1990) parallel distributed processing model, which places an emphasis on the relative strengths of pathways in the processing system, provides a satisfactory account for the number Stroop phenomenon and the performances differences across different writing scripts in bilingual subjects. The present thesis argues that when discriminability is incorporated, as a factor influencing comparison Stroop task performance, into Cohen et al.'s (1990) model, distance effects (both classic and reversed) and effects of familiarity could also be accounted for.

7.2.2 Numerical Magnitude Representation

The current thesis focuses on the processing and representation of numerical magnitudes. The series of experiments reported here have provided new insights into how numerical magnitudes are represented. In this section, three further aspects of numerical representation are considered: 1) whether numbers are mentally represented from left to right, 2) whether numbers are represented on a compressive mental number line, and 3) whether numbers evoke the same type of representation as other quantitative stimuli.

The first aspect of numerical representation being considered in the current section is whether numbers are mentally represented from left to right. Dehaene et al. (1993), in a series of experiments involving parity judgements, reported the SNARC effect – the preferential rightward response with large numbers and leftward response with small numbers. This effect, observed with both single digits and verbal numerals (Dehaene et al., 1993; see also Fias et al., 1996), has been interpreted as a spatial congruency between the response side (left or right) and the relative position of numerical magnitudes on the mental number line. Furthermore, based on the finding that the SNARC effect was not observed with non-numerical stimuli in a consonant-vowel letter classification task (Dehaene et al., 1993, Experiment 4), the authors concluded that the SNARC effect is specific to numbers, and not any stimuli obeying a fixed sequential order, such as letters. They attributed the effect to the semantic (magnitude), rather than the ordinal representation of numbers. However, Gevers et al. (2003) observed the SNARC effect with non-numerical stimuli, demonstrating that the effect is not specific to numerical stimuli, and suggesting that the effect merely reflects the ordinal rather than the semantic (magnitude) information.

Experiments 5a and 5b were designed to examine whether numbers are mentally represented from left to right. In these experiments, subjects were asked to judge if two numbers had the same parity or not. If numbers were mentally represented from left to right, then the stimulus-side effect should emerge in such a way that a faster mean reaction time would be observed when the smaller number was presented on the left than when presented on the right. However, this was not observed, thus providing no support for the idea of a left-to-right mental representation of numbers.

The second aspect of numerical representation being considered in the current section is whether numbers are mentally represented on a compressive number line. This idea was originally proposed to account for the numerical distance effect in the numerical comparison task – the time to compare a number pair could be explained by a logarithmic function of the numerical distance between them (Aiken & Williams, 1968; Moyer & Landauer, 1967; Restle, 1970). Dehaene and colleagues, in a series of papers, argued that numbers are converted to analogue representations in the form of a mental number line (Dehaene, 1992; Dehaene & Changeux, 1993; Dehaene et al., 1990; see also McCloskey, 1992; Restle, 1970), and that this number line is said to be compressive (i.e., non-linear), obeying the Weber-Fechner logarithmic law (Weber, 1834; Fechner, 1860). Accordingly, the subjective difference (Δ) between two numbers depends on their positions on the number line, so that the subjective difference between N and N + 1 decreases as N increases. This can be expressed by the equation:

$$\Delta = (N + 1) / N$$

However, the concept of analogue representations for numbers has been challenged by Zorzi and Butterworth (1999) who argued that numerals are mapped linearly on to magnitude representations (discrete numerosities) and that subjective distance (Δ) between the number N and N + 1 can be expressed by the following equation:

$$\Delta = \log \left(\frac{N+1}{N} + k \right)$$

where k is a constant. Zorzi and Butterworth (1999), in a computer simulation, demonstrated the compressive effect during number comparisons without assuming non-linear (compressive) numerical magnitude representations. According to these authors, the compressive effect should not be attributed to the representations of numerical magnitudes. Instead, they argued that the effect resulted from intrinsic non-linear interactions that occurred in the decision process (see Section 1.4.6 for details).

Experiment 6b examined two predictions based on the compressive number line hypothesis (Dehaene, 1992; Restle, 1970) which assumes that the distance between adjacent magnitude representations associated with numbers on the mental number line decreases as the numbers increase. Although support was observed for this hypothesis in that, for a given intra-stimulus distance, mean reaction time was significantly slower for a larger numerosity than smaller one (e.g., 2 vs. 2 2 2; the latter would have a smaller mental distance according to the compressive number line hypothesis and would thus produce more interference), this finding could also be explained in terms of a mere increase in difficulty due to the increase in the number of possible patterns which could be formed by a larger numerosity (Peterson & Simon, 2000). Furthermore, for a given numerosity, mean reaction times did not differ significantly between small and large numerical magnitudes (e.g., 1 1 and 3 3). This finding implies that the subjective distance between adjacent numerical representations does not increase with numerical magnitudes. This is inconsistent with the compressive mental line which would predict a slower reaction time for the latter where the mental distance is smaller. On the other hand, the finding can be taken as evidence supporting linear representations of numbers.

The third aspect of numerical representation being considered in the current section is whether numbers evoke the same type of representation as other quantitative stimuli.

The discoverers of the numerical distance effect, Moyer and Landauer (1967) explained numerical magnitude comparison as a process "in which the displayed numerals are converted to analogue magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of line." This view is also shared by other researchers (Dehaene, 1992; Dehaene & Changeux, 1993; Dehaene et al., 1990; see also McCloskey, 1992; Restle, 1970). However, the concept of analogue representations for numbers has been challenged by Zorzi and Butterworth (1999), who argued that numbers evoke discrete numerosity representations which follow the cardinality principle.

The neuroimaging data in Experiment 3b – enhanced activity was observed in several parietal regions (right inferior parietal lobule, right precuneus, and left superior parietal lobule) when processing numerical distance compared to processing physical distance – suggest that comparative judgements on discrete numerosity representations evoked by numbers call for higher processing requirements compared to those on analogue representations evoked by physical sizes (see Zorzi & Butterworth, 1999).

In summary, findings reported in the present thesis provide support for linear representations of numerical magnitudes, which are discrete, and hence distinct from continuous representations evoked by other quantitative stimuli, e.g., physical sizes. Although the literature on number research has come to the consensus that mental representations of numbers are directional, i.e., from left to right, current findings have provided no support for this notion.

7.2.3 Other Aspects of Number Semantics

The present thesis has focused primarily on numerical magnitude. Other aspects of number semantics need to be considered in order to gain a comprehensive understanding of numerical representation.

Existing models of numerical representation and processing have their limitations. For instance, Dehaene's (1992) triple-code model (Figure 1.2) consists of three types of mental numerical representations: an analogue magnitude code which is distinct from other number semantics, for example, parity information (visual Arabic code), and arithmetical facts (auditory verbal code). As discussed in the previous section, the model does not however distinguish between discrete and continuous/ analogue representations. Cipolotti and Butterworth's (1995) multiroute model (Figure 1.3), on the other hand, has defined elaborate semantic and asemantic routes for digit and verbal numeral processing, but lacks distinctions between different representations of number semantics.

As well as numerical magnitudes, numbers convey other information, including learnt facts such as parity (oddness and evenness), prime (number which is only divisible by 1 and itself), and arithmetical facts (addition, subtraction, multiplication, and division). It is therefore reasonable to propose distinctions for abstract internal representations of these semantic aspects of numbers. Perhaps within Cipolotti and Butterworth's (1995) multiroute model, the abstract internal representations may be dissociated into the following components: magnitude (which subdivides into discrete numerosity and continuous quantity), parity, prime, and arithmetical facts; these components interact with one another depending on task requirements.

7.3 Future Directions

All the experiments presented in the current thesis employed symbolic stimuli, i.e., digits and written verbal numerals. As reported by Girelli et al. (2000), the autonomy aspect of numerical magnitude processing would only emerge after children have understood the meanings of numbers (or numerical magnitudes) – children younger than 8 years old, who had not fully understood the association between numbers and their semantics (numerical magnitudes), did not exhibit the Stroop effect during physical comparisons. This has led to the question of whether non-symbolic quantity comparison would be age and knowledge dependent. Future research should concentrate on using non-symbolic stimuli, for example, using the comparison type Stroop paradigm, two displays of differently sized squares could be presented to children, who would be asked to select the display that contained either more squares (numerosity comparison task) or the larger total area taken up by the squares (area comparison task). The prediction is that young children who have not yet developed a comprehensive understanding of numbers and their semantics would still exhibit the Stroop effect with non-symbolic stimuli as would adults.

The comparison type Stroop paradigm with numbers has been used as a diagnostic tool for children with dyscalculia (Butterworth, 2003). If the aetiology of this developmental condition rests on a more general quantity processing deficit, then a non-symbolic version would also reveal abnormal performance patterns and could potentially be used with children at younger ages and provide an earlier diagnosis of dyscalculia.

References

Aiken, L. R., & Williams, E. M. (1968). Three variables related to reaction time to compare single digit numbers. *Perceptual and Motor Skills*, **27**, 199-206.

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition*, **24(5)**, 557-572.

Anderson, J. A., Spoehr, K. T., & Bennett, D. J. (1994). A study in numerical perversity: Teaching arithmetic to a neural network. In D. S. Levine & M. Aparicio (Eds.), *Neural networks for knowledge representation and inference*. Hillsdale, NJ: Erlbaum.

Andres, M., Davare, M., Pesenti, M., Oliver, E., & Seron, X. (2004). Number magnitude and grip aperture interaction. *Neuroreport*, **15**, 1-5.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, **13**, 263-308.

Ashbridge, E., Walsh, V., & Cowey, A. (1997). Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia*, **35(8)**, 1121-31.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75-106.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgements. In: G. H. Bower (Ed.), *The psychology of learning and motivation*, (Vol. 11, pp. 101-159). San Diego, CA: Academic Press.

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgements of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 435-447.

Barch, D.M., Braver, T.S., Nystrom, L.E., Forman, S.D., Noll, D.C., & Cohen, J.D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, **35**, 1373-1380.

Basso, G., Nichelli, P., Frassinetti, F., & di Pellegrino, G. (1996). Time perception in a neglected space. *Neuroreport*, **7**, 2111-2114.

Beauvois, M.-F., & Derouesne, J. (1979). Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery and Psychiatry*, **42**, 1115-1124.

Bench, C., Frith, C., Grasby, P., Friston, K., Paulesu, E., Frackowiak, R., & Dolan, R. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia*, **31**, 907-922.

Benson, D. F., & Denckla, M. B. (1969). Verbal paraphasia as source of calculation disturbance. *Archives of Neurology*, **21**, 96-102.

Berch, D. B., Foley, E. J., Hill, R. J., & McDonough Ryan, P. (1999). Extracting Parity and Magnitude from Arabic Numerals: Developmental Changes in Number Processing and Mental Representation. *Journal of Experimental Child Psychology*, **74**, 286-308.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia*, **17**, 467-472.

Biederman, I., & Tsao, Y.-C. (1979). On Processing Chinese Ideographs and English Words: Some Implications from Stroop-Test Results. *Cognitive Psychology*, **11**, 125-132.

Bjoertomt, O., Cowey, A., & Walsh, V. (2002). Spatial neglect in near and far space investigated by repetitive transcranial magnetic stimulation. *Brain*, **125**, 2012-22.

Boles, D. B. (1986). Hemispheric differences in the judgment of number. *Neuropsychologia*, **24**, 511-519.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, **108(3)**, 624-652.

Brown, S. W. (1997). Attentional resources in timing: interference effects in concurrent temporal and non-temporal working memory tasks. *Perception and Psychophysics*, **59**, 1118-1140.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: General*, **124**, 434-452.

Bunge, S.A., Dudukovic, N.M., Thomason, M.E., Vaidya, C.J., & Gabrieli, J.D.E. (2002). Development of frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron*, **33**, 301-311.

Bush, G., Frazier, J.A., Rauch, S.L., Seidman, L.J., Whalen, P.J., Jenike, M.A., Rosen, B.R., & Biederman, J. (1999). Anterior cingulate cortex dysfunction in attention deficit/hyperactivity disorder revealed by fMRI and the counting Stroop. *Biological Psychiatry*, **45**, 1542-1552.

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., Mclnerney, S. C., & Rauch, S. L. (1998). The Counting Stroop: An Interference Task Specialized for Functional Neuroimaging – Validation Study With Functional MRI. *Human Brain Mapping*, **6**, 270-282.

Butterworth, B. (2003). Dyscalculia Screener. nferNelson.

Butterworth, B., Cappelletti, M., & Kopelman, M. (2001). Category specificity in reading and writing: the case of number words. *Nature Neuroscience*, **4(8)**, 784-786.

Campbell, J. I. D. (1992). In defence of the encoding-complex approach: Reply to McCloskey, Macaruso and Whetstone. In: J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills* (pp. 539-556). Amsterdam: Elsevier Science.

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, **53**, 1-44.

Campbell, J. I. D., & Clark, J. M. (1988). An encoding complex view of cognitive number processing: Comment on McCloskey, Sokol and Goodman (1986). *Journal of Experimental Psychology: General*, **117**, 204-214.

Campbell, J. I. D., & Clark, J. M. (1992). Numerical cognition: An encoding-complex perspective. In: J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills* (pp. 457-491). Amsterdam: Elsevier Science.

Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case for single-patient studies. *Brain and Cognition*, **5**, 41-66.

Carey, S. (2001). Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language*, **16**, 37-55.

Carter, C.S., Beaver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, **280**, 747-749.

Carter, C., Mintun, M., & Cohen, J. (1995). Interference and facilitation effects during selective attention: An H_2 ^{15}O PET study of Stroop task performance. *Neuroimage*, **2**, 264-272.

Castelli, F., Glaser, D., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences*, in press.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, **11**, 63-65.

Cattell, J. M. (1902). The time of perception as a measure of differences in intensity. *Philosophische Studien*, **19**, 63-68.

Cheng, C. M., & Shih, S. I. (1988). The nature of lexical access in Chinese: Evidence from experiments on visual and phonological priming in lexical judgment. In I. M. Liu, H. C. Chen, & M. J. Chen (Eds.), *Cognitive aspects of the Chinese language* (pp. 1-14). Hong Kong, China: Asian Research Service.

Chochn, F., Cohen, L., van de Moortele, P. F., & Dehaene, S. (1999). Differential Contributions of the Left and Right Inferior Parietal Lobules to Number Processing. *Journal of Cognitive Neuroscience*, **11**, 617-630.

Cipolotti, L. (1993). Acquired disorders of numerical processing. *Unpublished Doctoral Thesis*. University of London.

Cipolotti, L. (1995). Multiple routes for reading words, why not numbers? Evidence from a case of Arabic numeral dyslexia. *Cognitive Neuropsychology*, **12**, 313-362.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired number transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, **124**, 375-90.

Cipolotti, L., Butterworth, B., & Warrington, E. K. (1995). Selective impairment in manipulating Arabic numerals. *Cortex*, **31**, 73-86.

Clapp, F. L. (1924). The number combinations: Their relative difficulty and frequency of their appearance in textbooks. *Research Bulletin*, **1**. Madison, WI: Bureau of Educational Research.

Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, **17**, 204-239.

Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control and schizophrenia: Recent development and current challenges. *Philosophical Transactions of the Royal Society B*, **351**, 1515-1527.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, **97**, 332-361.

Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In: C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance*, **XV**, 453-476. Cambridge, MA: MIT Press.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, **99**, 45-77.

Cohen, L., & Dehaene, S. (1991). Neglect dyslexia for numbers? A case report. *Cognitive Neuropsychology*, **8**, 39-58.

Cohen, L., & Dehaene, S. (1995). Number processing in pure alexia: The effect of hemispheric asymmetries and task demands. *Neurocase*, **1**, 121-137.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *NeuroImage*, **22**, 466-476.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., et al. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, **123**, 291-307.

Cohen, L., Dehaene, S., & Verstichel, P. (1994). Number words and number non-words. A case of deep dyslexia extending to Arabic numerals. *Brain*, **117**, 267-279.

Coltheart, M. (1978). Lexical access in simple reading tasks. In: G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

Coltheart, M. (1982). The psycholinguistic analysis of the acquired dyslexias: Some illustrations. *The Philosophical transactions of the Royal Society of London*, **B298**, 151-164.

Critchley, H. D., Tang, J., Glaser, D., Butterworth, B., & Dolan, R. J. (2005). Anterior cingulate activity during error and autonomic responses. *NeuroImage*, **27(4)**, 885-895.

Crosby, A.W. (1997). *The Measure of Reality.* New York: Cambridge University Press.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception and Psychophysics*, 45, 557-566.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, **44**, 1-42.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **21**, 314-326.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General*, **122**, 371-396.

Dehaene, S., & Changeux, J. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, **5**, 390-407.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia*, **29**, 1045-1074.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, **1**, 83-120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, **33**, 219-250.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 626-641.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, **43**, 1-29.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioural and brain-imaging evidence. *Science*, **284**, 970-3.

Delazer, M., & Butterworth, B. (1997). A dissociation of number meanings. *Cognitive Neuropsychology*, **14**, 613-636.

Deloche, G., & Seron, X. (1982a). From one to 1: an analysis of a transcoding process by means of neuropsychological data. *Cognition*, **12**, 119-149.

Deloche, G., & Seron, X. (1982b). From three to 3: A differential analysis of skills in transcoding quantities between patients with Broca's and Wernicke's aphasia. *Brain*, **105**, 719-733.

Deloche, G., & Seron, X. (1987). Numerical transcoding: a general production model: In: G. Deloche, & X. Seron (Eds.), *Mathematical disabilities: A cognitive neuropsychological perspective* (pp.137-170). Hillsdale, NJ: Lawrence Erlbaum.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., & Mazoyer, B. (1996). Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia*, **34(11)**, 1097-1106.

De Long, A. J. (1981). Phenomenological space-time: towards: an experimental relativity. *Science*, **213**, 681-683.

Derbyshire, S., Vogt, B., & Jones, A. (1998). Pain and Stroop interference tasks activate separate processing modules in anterior cingulate cortex. *Experimental Brain Research*, **118**, 52-60.

Dunbar, K. N., and MacLeod, C. M. (1984). A horse race of a different color: Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 622-639.

Duncan, E. M., & MacFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: an additive factors analysis. *Memory & Cognition*, **8**, 612-622.

Durston, S., Thomas, K.M., Yang, Y., Uluğ, A.M., Zimmerman, R.D., & Casey, B.J. (2002). A neural basis for the development of inhibitory control. *Developmental Science*, **5(4)**, F9-F16.

Dyer, F. N. (1971). Color-namimg interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, **10**, 297-302.

Effler, M. (1978). Colour-distant versus colour-congruent colour words, their influence on naming times in the Stroop-test. *Psychologische Beiträge*, **20**, 345-359. (From *Psychological Abstracts*, 1980, **64**, Abstract No. 2591).

Fairbank, B. A. (1969). Experiments on the temporal aspects of number perception. *Unpublished doctoral dissertation*. University of Arizona.

Fan, K. Y., Gao, J. Y., & Ao, X. P. (1984). Hanzi he pinyin wenzi de duyin guize [Pronunciation principles of Chinese characters and alphabetic writing scripts]. *Han zi gai ge [Chinese Character Reform]*, **3**, 23-27.

Fechner, G. T. (1860). *Elemente der Psychophysik* (Two Volumes). Breitkopf & Härtel: Leipzig.

Fias, W. (2001). Two routes for the processing of verbal numbers: evidence from the SNARC effect. *Psychological Research*, **65(4)**, 250-9.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*, **2**, 95-110.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal Representation of Symbolic and Nonsymbolic Magnitude. Journal of *Cognitive Neuroscience*, **15(1)**, 47-56.

Flowers, J. H., & Stoup, C. M. (1977). Selective attention between words, shapes and colors in speeded classification and vocalization tasks. *Memory & Cognition*, **5**, 299-307.

Flowers, J. H., Warner, J. L., & Polansky, M. L. (1979). Response and encoding factors in "ignoring" irrelevant information. *Memory & Cognition*, **7**, 86-94.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparisons of size and magnitude. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **10**, 442-453.

Frick, R. W. (1987). The homogeneity effect in counting. *Perception and Psychophysics*, **41(1)**, 8-16.

Fuson, K. C. (1988). *Children's Counting and Concepts of Number*. New York: Springer-Verlag.

Fuson, K.C. (1992). Relationships between counting and cardinality from age 2 to 8. In: J. Bideaud, C. Meljac, & J. P. Fisher (Eds), *Pathways to Number: Children's Developing Numerical Abilities*. Hillsdale, NJ: Lawrence Erlbaum.

Galfano, G., Rusconi, E., & Umiltà, C. (2003). Automatic activation of multiplication facts: Evidence from the nodes adjacent to the product. *The Quarterly Journal of Experimental Psychology A*, **56**, 31-63.

Gallistell, R. C., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, **44**, 43-74.

Gallistell, R. C., & Gelman, R. (2000). Non-verbal numerical cognition: from real to integers. *Trends in Cognitive Sciences*, **4**, 59-65.

Galton, F. (1880). Visualised numerals. *Nature*, **21**, 252-256.

Garavan, H., Ross, T.J., & Stein, E.A. (1999). Right hemispheric dominance of inhibitory control: an event-related fMRI study. *Proceedings of the National Academy of Sciences USA*, **96(14)**, 8301-8306.

George, M., Ketter, T., Parekh, P., Rosinsky, N., Ring, H., Casey, B., Trimble, M., Horwitz, B., Herscovitch, R., & Post, R. (1994). Regional brain activity when selecting a response despite interference: An H215O PET study of the Stroop and an emotional Stroop. *Human Brain Mapping*, **1**, 194-209.

Gelman, R., & Gallistel, C. R. (1978). The *Child's Understanding of Number.* Cambridge, MA: Harvard University Press.

Gelman, R., & Gallistel, C. R. (1986). The *Child's Understanding of Number.* Cambridge, MA: Harvard University Press. [1st edn 1978.]

Gelman, R. & Meck, E. (1983). Preschoolers counting: Principles before skill. *Cognition*, **13**, 343-59.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, **87**, B87-B95.

Gilbert, S., & Shallice, T. (2001). Task switching: A PDP Model. *Cognitive Psychology*, **44(3)**, 297-337.

Girelli, L. (1998). Accessing number meaning in adults and children. *A Thesis Submitted to the Faculty of Life Science of the University of London for the degree of Doctor of Philosophy.*

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, **76**, 104-122.

Glaser, M.O., & Glaser, W.R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 875-894.

Göbel, S., Rushworth, M. F. S., & Walsh, V. (2001a). rTMS disrupts the representation of small numbers in supramarginal gyrus [Abstract]. *NeuroImage*, **13**(6, Suppl. 1), 409.

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001b). The mental number line and the human angular gyrus. *NeuroImage*, **14(6)**, 1278-89.

Harbeson, M. M., Kennedy, R. S., & Bittner, A. C. (1982). The Stroop as a performance evaluation test for environmental research. *Journal of Psychology*, **111**, 223-233.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory and Cognition*, **10**, 389-395.

Hines, T. M. (1990). An odd effect: Lengthened reaction times for judgements about odd digits. *Memory & Cognition*, **18**, 40-46.

Hinrichs, J. V., & Yurko, D. S., & Hu, J. H. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 890-901.

Hung, D. L., & Tzeng, O. J. L. (1981). Orthographic variations and visual information processing. *Psychological Bulletin*, **90**(3), 377-414.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.

Kiehl, K.A., Kiddle, P.F., Hopfinger, J.B. (2000). Error processing and the rostral anterior cingulate: an event-related fMRI study. *Psychophysiology*, **37**, 216-223.

Kolers, P. A. (1968). Bilingualism and information processing. *Scientific American*, **218**, 78-86.

Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., and Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related fMRI. *Brain*, **122**, 981-991.

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., & Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *European Journal of Neuroscience*, **10**, 1209-1213.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

LeFevre, J., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition,* **16**, 45-53.

Leong, C. K. (1997). Paradigmatic analysis of Chinese word reading: Research findings and classroom practices. In: C. K. Leong & R. M. Joshi (Eds.), *Cross-language studies of learning to reading and spell: Phonological and orthographic processing* (pp. 379-417). Dordrecht/ Norwell, MA: Kluwer Academic.

Liu, A.-Y. (1973). Decrease in Stroop effect by reducing semantic interference. *Perceptual Motor Skills*, **37**, 263-265.

Logan, G. D. (1978). Attention in character-classification tasks: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General*, **107**, 32-63.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, **12**, 523-553.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Lories, G., Aubrun, A., & Seron, X. (1994). Lesioning McCloskey & Lindemann's (1992) MATHNET: The effect of damage location and amount. *Journal of Biological Systems*, **2**, 335-356.

Lovelace, E. A., & Snodgrass, R. D. (1971). Decision time for alphabetic order of letter pairs. *Journal of Experimental Psychology*, **88**, 258-264.

Macaruso, P., McCloskey, M., & Aliminosa, D. (1993). The functional architecture of the cognitive numerical-processing system: Evidence from a patient with multiple impairments. *Cognitive Neuropsychology*, **1**, 341-376.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109(2)**, 163-203.

Mandler, G., & Shebo, B. J. (1949). Subitizing: An Analysis of Its Component Processes. *Journal of Experimental Psychology: General*, **3(1)**, 1-22.

Marshall, J. C., & Halligan, P. W. (1989). When right goes left, An investigation of line bisection in a case of visual neglect. *Cortex*, **25**, 503-515.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, **2**, 175-199.

McCloskey, M. (1992). Cognitive mechanisms in number processing: Evidence from acquired dyscalculia. *Cognition*, **44**, 107-157.

McCloskey, M., Aliminosa, D., & Sokol, S. M. (1991). Facts, rules and procedures in normal calculation: Evidence from multiple single-patient studies of impaired arithmetic fact retrieval. *Brain and Cognition*, **17**, 154-203.

McCloskey, M., & Caramazza, A. (1987). Cognitive mechanisms in normal and impaired number processing. In G. Deloche and X. Seron (Eds.), *Mathematical Disabilities: A Cognitive Neuropsychological* Perspective. Hillsdale, N.J.: Lawrence Erlbaum Associates.

McCloskey, M., & Caramazza, A. (1988). Theory and methodology in cognitive neuropsychology: a response to our critics. *Cognitive Neuropsychology*, **5**, 583-623.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, **4**, 171-196.

McCloskey, M., & Lindemann, A. (1992). Mathnet: preliminary results from a distributed model of arithmetic fact retrieval. In: J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills* (pp. 365-409). Amsterdam: Elsevier Science.

McCloskey, M., Macaruso, P., & Whetstone, T. (1992). The functional architecture of numerical processing mechanisms: defending the modular view. In: J. I. D. Campbell (Ed.), *The Nature and Origins of Mathematical Skills* (pp. 493-538). Amsterdam: Elsevier Science.

McCloskey, M., Sokol, S. M., & Goodman, R.A. (1986). Cognitive Processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General*, **115**, 307-330.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, **21**, 627-645.

Menon, V., Adleman, N.E., White, C.D., Glover, G.H., & Reiss, A.L. (2001). *Human Brain Mapping*, **12**, 131-143.

Milham, M.P., Banich, M.T., & Barad, V. (2003). Competition for priority in processing increases prefrontal cortex's involvement in top-down control: An event-related fMRI study of the Stroop Task. *Cognitive Brain Research*, **17**, 212-222.

Milham, M.P., Banich, M.T., Webb, A., Barad, V., Cohen, N.J, Wszalek, T., & Kramer, A.F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, **12**, 467-473.

Milham, M.P., Erickson, K.I., Banich, M.R., Kramer, A.F., Webb, A., Wszalek, T., & Cohen, N.J. (2002). Attentional control in the aging brain: Insights from an fMRI study of the Stroop task. *Brain and Cognition*, **49**, 277-296.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: In search of partial output. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 273-296.

Miller, K., & Gelman, R. (1983). The child's representation of number: A multidimensional scaling analysis. *Child Development*, **54**, 1470-1479.

Mitchell, C. T., & Davies, R. (1987). The perception of time in scale model environments. *Perception*, **16**, 5-16.

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, **60**, 329-346.

Moyer, R. S. (1973). Comparing objects in memory – Evidence suggesting an internal psychophysics. *Perception and Psychophysics*, **13**, 180-184.

Moyer, R. S., & Landauer, T. K. (1967). The time required for numerical inequality. *Nature*, **215**, 1519-1520.

Moyer, R. S., & Landauer, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*, **1**, 167-168.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353-383.

Nelson, J.K., Reuter-Lorenz, P.A., Sylvester, C-Y. C., Jonides, J., & Smith, E.E. (2003). Dissociable neural mechanisms underlying response-based and familiarity-based conflict in working memory. *Proceedings of the National Academy of Sciences*, **100**, 11171-11175.

Nielsen, G. D. (1975). The locus and mechanism of the Stroop color word effect (Doctoral dissertation, University of Wisconsin – Madison, 1974). *Dissertation Abstracts International*, **35**, 5672-B.

Noël, M. P., & Seron, X. (1993). Arabic number reading deficit: a single-case study or when 236 is read (2306) and judged superior to 1258. Cognitive *Neuropsychology*, **10**, 317-339.

Noël, M. P., & Seron, X. (1995). Lexicalization errors in writing Arabic numerals. *Brain and Cognition*, **29**, 151-179.

Noël, M. P., & Seron, X. (1997). On the existence of an intermediate semantic representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23**, 697-720.

Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *The Quarterly Journal of Experimental Psychology*, **57A(5)**, 835-863.

Nürk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, **82**, B25-B33.

Ogura, C. (1980). Formation of learning-set in the task of Stroop color-word test. *Journal of Child Development*, **16**, 30-36.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition*, **3**, 635-647.

Paivio, A. (1978). Comparisons of mental clocks. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 61-71.

Pansky, A., & Algom, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception and Performance*, **25(1)**, 39-58.

Pardo, J., Pardo, P., Janer, K., & Raichle, M. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Proceedings of the National Academy of Sciences USA*, **87(1)**, 256-259.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, **91**, 191-205.

Parry, A. M., Scott, R. B., Palace, J., Smith, S., & Matthews, P. M. (2003). Potentially adaptive functional changes in cognitive processing for patients with multiple sclerosis and their acute modulation by rivastigmine. *Brain*, **126(12)**, 2750-60.

Passingham, R.E. (1996). Functional specialisation of the supplementary motor area in monkeys and humans. In: H.O. Luders (Ed.), *Advances in Neurology*, vol. **70**: Supplementary Motor Area, pp. 105-116. Lippingcott-Raven, Philadelphia.

Pavese, A., & Umiltà, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance*, **24(5)**, 1535-1545.

Pavese, A., & Umiltà, C. (1999). Further evidence on the effects of symbolic distance on Stroop-like interference. *Psychological Research*, **62**, 62-71.

Perfetti, C. A., & Zhang, S. (1991). Phonological processes in reading Chinese characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 633-643.

Perfetti, C. A., & Zhang, S. (1995). Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 24-33.

Perfetti, C. A., Zhang, S., & Berent, I. (1992). Reading in English and Chinese: Evidence for a "universal" phonological principle. In: R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 227-248). Amsterdam: North-Holland.

Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuroanatomical Substrates of Arabic Number Processing, Numerical Comparison, and

Simple Addition: A PET Study. *Journal of Cognitive Science*, **12(3)**, 461-479.

Peterson, S. A., & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science*, **24(1)**, 93-122.

Phaf, R. H., Van der Heijder, A. H. C., & Huston, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, **22**, 273-341.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. (2002). Are Subitizing and Counting Implemented as Separate or Functionally Overlapping Processes? *NeuroImage*, **15**, 435-446.

Picard, N., & Strick, P.L. (1996). Motor areas of the medial wall: A review of their location and functional activation. *Cerebral Cortex*, **6**, 342-353.

Pinel, P., Dehaene, S., Rivière, D., & Le Bihan, D. (2001). Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. *NeuroImage*, **14**, 1013-1026.

Pinel, P., Le Clec'H, G., van de Moortele, P.F., Naccache, L., Le Bihan, D., & Dehaene, S. (1999). Event-related fMRI analysis of the cerebral circuit for number comparison. *NeuroReport*, **10**, 1473-1479.

Pinel, P., Piazza, M., Le Bihan, D., and Dehaene, S. (2004). Distributed and Overlapping Cerebral Representations of Number, Size, and Luminance during Comparative Judgments. *Neuron*, **41**, 983-993.

Piazza, M., Mechelli, A., Butterworth, B., & Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage*, **15(2)**, 435-46.

Posner, M., & DiGirolamo, G. (1998). Executive attention: Conflict, target detection, and cognitive control. In: R. Parasuraman (Ed.), *The Attentive Brain*, pp. 401-423. Cambridge, MA: MIT Press.

Posner, M. I., & Snyder, C.R.R. (1975). Attention and cognitive control. In: R. L. Solso (Ed.), *Information processing cognition: The Loyola Symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, **8**, 295-301.

Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, **19**, 473-481.

Rao, S. M., Mayer, A.R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nature Neuroscience*, **4(3)**, 317-23.

Ratinckx, E., Brysbaert, M., & Fias, W. (In Press). Naming two-digit Arabic numerals: Evidence from masked priming studies. *Journal of Experimental Psychology: Human Perception and Performance.*

Regan, J. E. (1977). Automatic processing (Doctoral dissertation, University of California, Berkeley, 1977). *Dissertation Abstracts International*, **39**, 1018-B.

Reisberg, D., Baron, J., and Kemler, D. G. (1980). Overcoming Stroop interference: The effects of practice on distractor potency. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 140-150.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology: General*, **91**, 191-205.

Reynvoet, B., & Brysbaert, M. (1999). Single-digit and two-digit numerals address the same semantic number line. *Cognition*, **72**, 191-201.

Reynvoet, B., & Brysbaert, M. (2004). Cross-notation number priming investigated at different stimulus onset asynchronies in parity and naming tasks. *Experimental Psychology*, **51**, 81-90.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *Quarterly Journal of Experimental Psychology A*, **55**, 1127-1139.

Roe, W. T, Wilsoncraft, W. E., & Griffiths, R. S. (1980). Effects of motor and verbal practice on the Stroop task. *Perceptual and Motor Skills*, **50**, 647-650.

Rossetti, Y., Jacquin-Courtois, S., Rode, G., Ota, H., Michel, C., & Boisson, D. (2004). Does Action Make the Link Between Number and Space Representation? Visuo-Manual Adaptation Improves Number Bisection in Unilateral Neglect. *Psychological Science*, **15(6)**, 426-430.

Rubia, K., Overmeyer, S.O., Taylor, E., Brammer, M., Williams, S., Simmons, A., Andrew, C., & Bullmore, E.T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher order motor control: A study with functional MRI. *American Journal of Psychiatry*, **156**, 891-896.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology*, **81**, 74-92.

Rusconi, E., Walsh, V., & Butterworth, B. (2005). Dexterity of numbers: rTMS over left angular gyrus disrupts finger gnosis and number processing. *Neuropsychologia*, **43**, 1609-1624.

Rushworth, M. F., Ellison, A., & Walsh, V. (2001). Complementary localization and lateralization of orienting and motor attention. *Nature Neuroscience*, **4(6)**, 656-61.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgements of numerical inequality. *Child Development*, **48**, 630-633.

Seron, X., & Deloche, G. (1983). From 4 to four: A supplement to "From three to 3". *Brain*, **106**, 735-744.

Seron, X., & Deloche, G. (1984). From 2 to two: An analysis of a transcoding process by means of neuropsychological evidence. *Journal of Psycholinguistic Research*, **13**, 215-235.

Seron, X., Pesenti, M., Noël, M.-P., Deloche, G., & Cornet, J.-A. (1992). Images of numbers, or "When 98 is upper left and 6 sky blue". *Cognition*, 44, 159-96.

Shallice, T., Warrington, E. K., & McCarthy, R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology*, **35A**, 111-138.

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, **7**, 82-138.

Shiffrin, R.M., & Shneider W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and general theory. *Psychological review*, **84**, 127-190.

Shor, R. E. (1971). Symbolic processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology*, **85**, 187-205.

Shor, R. E., Hatch, R. P., Hudson, L. J., Landrigan, D. T., & Shaffer, H. J. (1972). Effects of practice on a Stroop-like spatial directions task. *Journal of Experimental Psychology*, **94**, 168-172.

Shrager, J., Klahr, D., & Chase, W. (1982). Segmentation and quantification of random dot patterns. Paper presented at the 23rd annual meeting of the Psychonomics Society.

Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, **33(3)**, 475-87.

Singer, H. D., & Low, A. A. (1933). Acalculia (Henschen): A clinical study. *Archives of Neurology and Psychiatry*, **29**, 476-498.

Sokol, S., Goodman-Shulman, R., & McCloskey, M. (1989). In defence of a modular architecture for the number-processing system: Reply to Campbell and Clark. *Journal of Experimental Psychology: General*, **118**, 105-110.

Sokol, S., McCloskey, M. (1991). Cognitive mechanisms in calculation. In: R. Sternberg and P. A. Frensch (Eds.), *Complex Problem Solving: Principles and Mechanisms* (pp. 85-116). Hillsdale, NJ: Erlbaum.

Sokol, S., McCloskey, M., Cohen, N. J., & Aliminosa, D. (1991). Cognitive representations and processes in arithmetic: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**, 355-376.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, **13**, 89-103.

Takahashi, A., & Green, D. (1983). Numerical judgments with Kanji and Kana. *Neuropsychologia*, **21(3)**, 259-263.

Tan, L. H., Hoosain, R., & Peng, D.-L. (1995). Role of early presemantic phonological code in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 43-54.

Tan, L. H., Hoosain, R., & Siok, W. W. T. (1996). The activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 865-882.

Tan, L. H., & Peng, D.-L. (1991). Visual recognition processes of Chinese characters. *Acta Psychologica Sinica*, **3**, 272-278.

Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese Character Recognition: Does Phonological Information Mediate Access to Meaning? *Journal of Memory and Language*, **37**, 41-57.

Taylor, S., Kornblum, S., Minoshima, S., Oliver, L., & Koeppe, R. (1994). Changes in medial cortical blood flow with a stimulus-response compatibility task. *Neuropsychologia*, **32**, 249-255.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18(1)**, 166-179.

Tzeng, O. J. L., & Wang, W. S.-Y. (1983). The first two R's. *American Scientist*, **71**, 238-243.

van Veen, V., Cohen, J.D., Botvinick, M.M., Stenger, V.A., & Carter, C.S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage*, **14**, 1302-1308.

Vaid, J. (1985). Numerical size comparisons in a phonologically transparent script. *Perception & Psychophysics*, **37**, 592-595.

Verguts, T., Fias, W., & Steven, M. (2005). A model of exact small-number representation. *Psychonomic Bulletin & Review*, **12(1)**, 66-80.

Viscuso, S. R., Anderson, J. A., & Spoehr, K. T. (1989). Representing simple arithmetic in neural networks. In G. Tiberghien (Ed.), *Advances in cognitive science* (Vol. 2, pp. 141-164). Chichester, UK. Ellis Horwood.

Vuilleumier, P., Ortigue, S., & Brugger, P. (2004). The Number Space and Neglect. *Cortex*, **40**, 399-410.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, **7(11)**, 483-488.

Walsh, V., & Pasual-Leone, A. (2003). *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind.* MIT Press.

Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae.* C. F. Köhler: Leipzig.

Weissman, D. H., Giesbrecht, B., Song, A.W., Mangun, G. R., & Woldorff, M. G. (2003). Conflict monitoring in the human anterior cingulate cortex during selective attention to global and local object features. *NeuroImage*, **19**, 1361-1368.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. Ergonomics, 3, 189-230.

Whalen, J., & Morelli, F. (2002). Cognitive neuroscience of number. Poster presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

White, B. W. (1978). Interference proneness and the ability to shift attention in old age (Doctoral dissertation, University of Notre Dame, 1978). *Dissertation Abstracts International*, **39**, 2549-B.

Willmes, K., & Iversen, W. (1995). On the Internal Representation of Number Parity. Paper presented at the Spring Annual Meeting of the British Neuropsychological Society, London.

Windes, J. D. (1968). Reaction time for numerical coding and naming of numerals. *Journal of Experimental Psychology*, **78**, 318-322.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, 36, 155-93.

Yin, W. G., & Butterworth, B. (1992). Deep and surface dyslexia in Chinese. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese. Advances in psychology*, **90** (pp. 349-366). Amsterdam, Netherlands: North-Holland.

Yin, W. G., & Butterworth, B. (1998). Chinese pure alexia. *Aphasiology*, **12(1)**, 65-76.

Yurko, D. S., & Hinrichs, J. V. (1978). Judgement of numerical inequality: Size-value congruity, Paper presented at the Midwestern Psychological Association, Chicago, IL.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental process: a case study of arithmetic. *Journal of Experimental Psychology: General*, **115**, 118-130.

Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus-stimulus and stimulus-response compatibility. *Cognitive Psychology*, **38**, 386-432.

Zhou, Y. (1978). To what degree are the "phonetics" of present-day Chinese characters still phonetic? *Zhongguo Yuwen*, **146**, 172-177.

Zimmer, K. (1964). Affixed negation in English and other languages: An investigation of restricted productivity. *Word*, **20**, 2, Monograph No. 5.

Zorzi , M., & Butterworth, B. (1999). A Computational Model of Number Comparison. In: M. Hahn, & S. C. Stoness (Eds.), *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (pp. 772-777). Lawrence Erlbaum Associates, Publishers.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature*, **417**, 138-139.

Zorzi, M., Stoianov, I., Priftis, K., & Umiltà, C. (2003). Semantic priming with the numerosity representations: Connectionist simulation. Presented at the Twenty-first European Workshop on Cognitive Neuropsychology, Bressanone, Italy.

Zysset, S., Müller, K., Lohmann, G., & von Cramon, D. Y. (2001). Color-word matching Stroop task: Separating interference and response conflict. *NeuroImage*, **13**, 29-36.